

Topologia e Geometria degli Spazi Semantici

Fondamenti degli Spazi Vettoriali Semantici e Metriche di Similarità

Introduzione: La Geometrizzazione del Significato

La rappresentazione matematica del significato linguistico attraverso strutture geometriche costituisce una delle conquiste più significative della linguistica computazionale contemporanea. L'idea fondamentale che sottende questa prospettiva è che le relazioni semantiche tra le parole possano essere catturate attraverso relazioni spaziali in opportuni spazi vettoriali, dove la prossimità geometrica riflette l'affinità di significato. Questa intuizione, che affonda le sue radici nella linguistica distribuzionale di Harris e nella sua celebre ipotesi secondo cui "parole che occorrono in contesti simili hanno significati simili", ha trovato una formalizzazione matematica rigorosa attraverso la teoria degli spazi vettoriali e la geometria differenziale.

Il paradigma degli spazi semantici rappresenta una sintesi straordinaria tra la linguistica teorica, la matematica pura e l'apprendimento automatico, dove concetti astratti come il significato, la sinonimia, l'antonimia e le relazioni semantiche complesse vengono mappati in strutture geometriche misurabili, calcolabili e manipolabili algebricamente. La potenza di questo approccio risiede non soltanto nella sua capacità di catturare regolarità statistiche nei corpora linguistici, ma anche nel fornire un framework unificato per comprendere fenomeni linguistici complessi come la composizionalità semantica, la polisemia, l'ambiguità lessicale e l'evoluzione diacronica del significato.

Spazi Vettoriali Semantici: Fondamenti Matematici

Uno spazio vettoriale semantico può essere formalmente definito come una tupla $S = (V, d, \phi)$, dove V è uno spazio vettoriale di dimensione finita su un campo K (tipicamente i numeri reali $\mathbb{R}$), d è una funzione di distanza o una metrica che quantifica la dissimilarità semantica, e ϕ è una funzione di embedding che mappa elementi di un vocabolario W in vettori di V . La costruzione di tale spazio richiede scelte architettoniche fondamentali che influenzano profondamente le proprietà geometriche e topologiche dello spazio risultante.

La dimensionalità dello spazio vettoriale costituisce un parametro cruciale. Tradizionalmente, gli spazi semantici sono stati costruiti con dimensioni che vanno da alcune decine a diverse centinaia di dimensioni, con valori tipici compresi tra 50 e 300 per i modelli di word embedding classici. La scelta della dimensionalità riflette un compromesso fondamentale tra la capacità espressiva del modello e la sua trattabilità computazionale. Spazi di dimensione troppo bassa non riescono a catturare la ricchezza e la complessità delle relazioni semantiche presenti nel linguaggio naturale, mentre spazi di dimensione eccessivamente

elevata soffrono della maledizione della dimensionalità, dove la concentrazione della misura rende meno significative le nozioni di prossimità e distanza.

La funzione di embedding $\phi: W \rightarrow \mathbb{R}^n$ rappresenta il cuore della costruzione dello spazio semantico. Esistono fondamentalmente due approcci principali per la definizione di tale funzione. Il primo approccio, di natura matriciale e basato sulla fattorizzazione, costruisce inizialmente una matrice di co-occorrenza C dove l'elemento C_{ij} quantifica la frequenza con cui la parola w_i appare nel contesto della parola w_j . Questa matrice, tipicamente di dimensioni $|W| \times |W|$ o $|W| \times |C|$ dove C rappresenta l'insieme dei contesti, viene poi fattorizzata attraverso tecniche come la decomposizione ai valori singolari (SVD) o la fattorizzazione matriciale non negativa (NMF), ottenendo una rappresentazione di dimensionalità ridotta che preserva le informazioni semantiche più rilevanti.

Il secondo approccio, di natura predittiva e basato su reti neurali, costruisce la funzione di embedding attraverso l'ottimizzazione di un obiettivo predittivo. I modelli Word2Vec, introdotti da Mikolov e collaboratori nel 2013, esemplificano questo paradigma. In particolare, il modello Skip-gram definisce la funzione obiettivo come la massimizzazione della probabilità logaritmica media di predire le parole di contesto dato una parola centrale. Formalmente, dato un corpus di parole $w_1, w_2, \dots, w_T$, l'obiettivo è massimizzare la funzione $J(\theta) = (1/T) \sum_{t=1}^T \sum_{-c \leq j \leq c, j \neq 0} \log p(w_{t+j} | w_t)$, dove c rappresenta la dimensione della finestra di contesto e $p(w_O | w_I)$ è definito attraverso la funzione softmax come $p(w_O | w_I) = \exp(v'_w w_O) / \sum_{w=1}^W \exp(v'_w w)$, dove v_w e v'_w rappresentano rispettivamente le rappresentazioni vettoriali di input e output della parola w .

L'architettura CBOW (Continuous Bag of Words), complementare allo Skip-gram, inverte la direzione della predizione cercando di prevedere la parola centrale dato il contesto circostante. La funzione obiettivo diventa $J(\theta) = (1/T) \sum_{t=1}^T \log p(w_t | w_{t-c}, \dots, w_{t-1}, w_{t+1}, \dots, w_{t+c})$, dove la rappresentazione del contesto è tipicamente ottenuta come media aritmetica dei vettori delle parole di contesto. Questa differenza architetturale ha conseguenze importanti sulle proprietà geometriche degli spazi risultanti, con CBOW che tende a produrre rappresentazioni più uniformi e Skip-gram che cattura meglio le relazioni sintattiche e le sottili sfumature semantiche.

Il modello GloVe (Global Vectors), proposto da Pennington, Socher e Manning nel 2014, rappresenta una sintesi elegante tra l'approccio matriciale e quello predittivo. GloVe definisce la funzione obiettivo come $J = \sum_{i,j=1}^V f(X_{ij})(v_i^T \tilde{v}_j + b_i + \tilde{b}_j - \log X_{ij})^2$, dove X_{ij} rappresenta il numero di volte che la parola j appare nel contesto della parola i , f è una funzione di pesatura che riduce l'influenza delle co-occorrenze troppo frequenti o troppo rare, v_i e $\tilde{v}_j$ sono i vettori di parola e di contesto, e b_i e $\tilde{b}_j$ sono termini di bias. La funzione di pesatura tipicamente utilizzata è $f(x) = (x/x_{\max})^\alpha$ se $x < x_{\max}$, e $f(x) = 1$ altrimenti, con $x_{\max}$ tipicamente fissato a 100 e $\alpha = 0,75$.

Metriche di Similarità Lessicale: Fondamenti Teorici

La scelta della metrica di similarità costituisce un aspetto fondamentale nella caratterizzazione degli spazi semantici, poiché determina quali configurazioni geometriche vengono considerate semanticamente equivalenti e quali proprietà topologiche lo spazio eredita. La metrica più comunemente utilizzata negli spazi vettoriali semantici è la similarità

del coseno, definita come $\cos(\theta) = (u \cdot v) / (\|u\| \|v\|) = \sum_{i=1}^n u_i v_i / (\sqrt{\sum_{i=1}^n u_i^2} \sqrt{\sum_{i=1}^n v_i^2})$, dove u e v sono due vettori di parola e θ rappresenta l'angolo compreso tra di essi. La similarità del coseno assume valori nell'intervallo $[-1, 1]$, dove 1 indica perfetta similarità direzionale, 0 indica ortogonalità (assenza di relazione lineare), e -1 indica opposizione direzionale.

La scelta della similarità del coseno rispetto ad altre metriche, come la distanza euclidea, è motivata da considerazioni sia teoriche che empiriche. Dal punto di vista teorico, la similarità del coseno è invariante rispetto alla normalizzazione in scala dei vettori, catturando così la direzione semantica piuttosto che la magnitudine. Questa proprietà è particolarmente rilevante negli spazi semantici, dove la magnitudine di un vettore può riflettere la frequenza di occorrenza di una parola piuttosto che la sua semantica intrinseca. Dal punto di vista empirico, la similarità del coseno ha dimostrato prestazioni superiori in una vasta gamma di compiti di valutazione della similarità semantica.

La distanza euclidea, definita come $d(u, v) = \|u - v\| = \sqrt{\sum_{i=1}^n (u_i - v_i)^2}$, rappresenta un'alternativa naturale che induce una struttura metrica propriamente detta sullo spazio vettoriale. Mentre la similarità del coseno misura l'angolo tra vettori, la distanza euclidea misura la loro separazione assoluta nello spazio. Queste due metriche non sono equivalenti: due vettori possono avere alta similarità del coseno ma grande distanza euclidea se hanno magnitudini molto diverse. Tuttavia, su vettori normalizzati alla lunghezza unitaria, esiste una relazione semplice tra le due metriche: $d^2(u, v) = 2(1 - \cos(\theta))$, il che significa che su sfere unitarie la distanza euclidea è una trasformazione monotona della similarità del coseno.

La distanza di Manhattan o L_1 , definita come $d_1(u, v) = \sum_{i=1}^n |u_i - v_i|$, e più in generale le distanze di Minkowski L_p , definite come $d_p(u, v) = (\sum_{i=1}^n |u_i - v_i|^p)^{1/p}$, forniscono una famiglia di metriche che generalizzano la distanza euclidea (caso $p = 2$). La scelta del parametro p influenza la sensibilità della metrica alle differenze nelle singole dimensioni: valori piccoli di p (come $p = 1$) rendono la metrica più sensibile a molte piccole differenze, mentre valori grandi di p enfatizzano le dimensioni con le differenze maggiori. Nel limite $p \rightarrow \infty$, si ottiene la distanza di Chebyshev $d_\infty(u, v) = \max_i |u_i - v_i|$.

La divergenza di Kullback-Leibler, sebbene non sia una metrica in senso stretto poiché non soddisfa la simmetria né la disuguaglianza triangolare, rappresenta una misura di dissimilarità importante quando i vettori sono interpretati come distribuzioni di probabilità. Definita come $D_{KL}(P||Q) = \sum_i P(i) \log(P(i)/Q(i))$, la divergenza KL misura l'informazione perduta quando si usa la distribuzione Q per approssimare la distribuzione P . Nelle applicazioni semantiche, questa divergenza è particolarmente rilevante quando si lavora con rappresentazioni probabilistiche dei significati o con modelli di linguaggio.

La distanza di Wasserstein o Earth Mover's Distance rappresenta una metrica più sofisticata che trova applicazione quando si desidera confrontare distribuzioni di probabilità tenendo conto della struttura geometrica dello spazio sottostante. Definita come $W(P, Q) = \inf_{\gamma \in \Gamma(P, Q)} \int \|x - y\| d\gamma(x, y)$, dove $\Gamma(P, Q)$ rappresenta l'insieme di tutte le distribuzioni congiunte con marginali P e Q , la distanza di Wasserstein può essere interpretata come il costo minimo di trasportare la massa di probabilità da una distribuzione all'altra. Questa metrica ha trovato applicazioni recenti nella costruzione di spazi semantici che preservano strutture gerarchiche.

Proprietà Topologiche degli Spazi Semantici

Gli spazi vettoriali semantici ereditano naturalmente una topologia indotta dalla metrica scelta. Nel caso della similarità del coseno su vettori normalizzati, lo spazio semantico può essere visto come un sottoinsieme della sfera unitaria $S^{n-1} = \{x \in \mathbb{R}^n : \|x\| = 1\}$ in $\mathbb{R}^n$, dotata della topologia indotta dalla distanza geodetica sulla sfera. Questa struttura sferica ha conseguenze importanti per la comprensione delle proprietà globali dello spazio semantico.

La compattezza della sfera unitaria implica che lo spazio semantico normalizzato sia compatto, il che garantisce, per il teorema di Bolzano-Weierstrass, che ogni sequenza infinita di vettori di parole contenga una sottosequenza convergente. Questa proprietà ha implicazioni pratiche per la stabilità degli algoritmi di clustering e per l'esistenza di configurazioni ottimali in problemi di ottimizzazione semantica. La compattezza, inoltre, implica che ogni ricoprimento aperto dello spazio ammetta un sottoricoprimento finito, una proprietà che può essere sfruttata nella costruzione di rappresentazioni gerarchiche compatte del lessico.

La connessione della sfera unitaria (nel senso che non può essere partizionata in due sottoinsiemi aperti non vuoti e disgiunti) riflette l'intuizione che lo spazio semantico forma un continuum di significati, dove è possibile muoversi gradualmente da un significato a un altro attraverso transizioni continue. Questa proprietà contrasta con concezioni discrete o categoriali del significato e supporta modelli di somiglianza semantica basati su gradienti piuttosto che su confini netti.

La curvatura costante positiva della sfera ha implicazioni per la distribuzione delle distanze nello spazio semantico. In particolare, su una sfera di raggio unitario in dimensione alta, la maggior parte dei punti casuali si concentra in una fascia equatoriale, un fenomeno noto come concentrazione della misura. Questo comportamento è rilevante per comprendere la distribuzione delle similarità semantiche: in uno spazio di dimensione molto alta, la maggior parte delle coppie di parole tende ad avere similarità del coseno prossime a zero, rendendo le alte similarità particolarmente informative.

L'omogeneità della sfera (il fatto che ogni punto della sfera può essere mappato in ogni altro punto attraverso un'isometria) implica che non esistono punti o regioni privilegiate nello spazio semantico puramente dal punto di vista geometrico. Ogni parola occupa una posizione geometricamente equivalente a ogni altra. Tuttavia, questa simmetria geometrica viene rotta dalla distribuzione effettiva delle parole nello spazio, che riflette le asimmetrie statistiche del linguaggio naturale.

Struttura Locale e Varietà Semantiche

A un livello locale, lo spazio semantico può essere analizzato attraverso la lente della teoria delle varietà differenziabili. Intorno a ogni punto (vettore di parola), lo spazio appare localmente euclideo, e possiamo definire strutture differenziali che permettono di studiare le variazioni infinitesimali del significato. Questa prospettiva è particolarmente fruttuosa quando si considerano famiglie parametriche di rappresentazioni semantiche, come quelle che emergono durante il training di modelli neurali.

Data una parola w con vettore associato $v \in \mathbb{R}^n$, lo spazio tangente $T_v S^{n-1}$ alla sfera nel punto v è definito come l'insieme dei vettori ortogonali a v : $T_v S^{n-1} = \{u \in \mathbb{R}^n : u \cdot v = 0\}$. Questo spazio tangente, di dimensione $n-1$, rappresenta lo spazio delle direzioni di variazione infinitesimale del significato accessibili da w . Le operazioni di composizione semantica, che esploreremo in seguito, possono essere interpretate come campi vettoriali sullo spazio semantico, che assegnano a ogni punto una direzione di variazione nel corrispondente spazio tangente.

La metrica riemanniana indotta sulla sfera dalla metrica euclidea ambiente permette di definire nozioni di lunghezza di curve e di geodetica. Le geodetiche sulla sfera sono archi di cerchio massimo, e rappresentano i percorsi di lunghezza minima tra due punti. Nel contesto semantico, una geodetica tra due parole può essere interpretata come un percorso di transizione semantica ottimale, che passa attraverso significati intermedi in modo graduale e continuo. La lunghezza di una geodetica, data dall'angolo θ compreso tra i due vettori, fornisce una misura intrinseca della distanza semantica che rispetta la geometria della varietà.

La curvatura gaussiana della sfera, costante e positiva, influenza il comportamento delle geodetiche: due geodetiche inizialmente parallele convergono necessariamente, riflettendo il fatto che percorsi semantici inizialmente indipendenti possono inevitabilmente intersecarsi in uno spazio di curvatura positiva. Questo fenomeno può essere messo in relazione con l'osservazione linguistica che campi semantici inizialmente distinti possono convergere attraverso processi metaforici o metonimici.

Clustering e Struttura Categoriale

La struttura categoriale del lessico emerge naturalmente dall'analisi della distribuzione dei vettori nello spazio semantico. Parole appartenenti alla stessa categoria semantica o sintattica tendono a formare cluster, ovvero regioni dello spazio caratterizzate da alta densità. L'identificazione di tali cluster può essere effettuata attraverso algoritmi di clustering come k-means, DBSCAN, clustering gerarchico agglomerativo, o metodi più sofisticati basati su modelli di mistura gaussiana.

L'algoritmo k-means partiziona lo spazio in k regioni di Voronoi, ciascuna associata a un centroide, minimizzando la funzione obiettivo $J = \sum_{i=1}^k \sum_{x \in C_i} \|x - \mu_i\|^2$, dove C_i rappresenta l' i -esimo cluster e μ_i il suo centroide. Nell'ambito semantico, i centroidi possono essere interpretati come prototipi categoriali, ovvero rappresentazioni ideali delle categorie semantiche. Tuttavia, l'algoritmo k-means assume cluster di forma sferica e richiede la specificazione a priori del numero di cluster, limitazioni che possono essere problematiche per la struttura complessa e annidata delle categorie semantiche.

DBSCAN (Density-Based Spatial Clustering of Applications with Noise) offre un'alternativa che identifica cluster basandosi sulla densità locale dei punti, senza richiedere la specificazione del numero di cluster. Un punto x è considerato un core point se la sua ϵ -sfera contiene almeno minPts punti. I cluster sono formati da core points connessi tra loro, e possono assumere forme arbitrarie. Questo approccio è particolarmente adatto per identificare campi semantici di dimensioni variabili e può identificare outlier, ovvero parole che non appartengono chiaramente a nessuna categoria dominante.

Il clustering gerarchico agglomerativo costruisce una gerarchia di cluster attraverso un processo bottom-up che parte da cluster singleton e progressivamente fonde cluster vicini. Questa gerarchia può essere rappresentata attraverso un dendrogramma, che visualizza le relazioni di inclusione tra cluster a diversi livelli di granularità. La scelta del criterio di linkage (single, complete, average, Ward) influenza la struttura della gerarchia risultante. Il criterio di Ward, che minimizza la varianza totale all'interno dei cluster, tende a produrre cluster compatti e ben bilanciati, risultando particolarmente adatto per categorie semantiche omogenee.

Analisi delle Relazioni Semantiche attraverso l'Algebra Vettoriale

Una delle scoperte più sorprendenti nell'ambito degli spazi semantici è che molte relazioni semantiche possono essere catturate attraverso semplici operazioni vettoriali. La celebre relazione vettoriale $v(re) - v(uomo) + v(donna) \approx v(regina)$, scoperta nel contesto di Word2Vec, illustra come relazioni analogiche possano essere codificate come traslazioni nello spazio vettoriale. Questa proprietà, nota come parallelogramma semantico, suggerisce che le relazioni semantiche come "è il genere femminile di" o "è la capitale di" corrispondano a direzioni costanti nello spazio.

Formalmente, una relazione semantica R può essere rappresentata come un vettore di offset r tale che se (a, b) soddisfa la relazione R , allora $v(b) \approx v(a) + r$. La consistenza di questa rappresentazione può essere valutata attraverso compiti di completamento di analogie della forma "a è a b come c è a d", cercando la parola d che massimizza la similarità del coseno con $v(b) - v(a) + v(c)$. L'accuratezza in questi compiti fornisce una misura della regolarità geometrica dello spazio semantico e della sua capacità di catturare relazioni semantiche composizionali.

Tuttavia, la rappresentazione lineare delle relazioni semantiche presenta limitazioni. Non tutte le relazioni semantiche sono catturabili come semplici traslazioni, e la qualità delle analogie varia significativamente a seconda del tipo di relazione considerata. Le relazioni sintattiche (come il plurale, i tempi verbali) tendono a essere rappresentate più consistentemente rispetto alle relazioni semantiche pure. Questo suggerisce che mentre la struttura locale dello spazio cattura bene certe regolarità linguistiche, la sua struttura globale potrebbe richiedere geometrie più complesse.

Estensione agli Spazi Contestualizzati

I modelli di embedding contestualizzati, come ELMo, GPT e BERT, rappresentano un'evoluzione significativa rispetto agli embedding statici, assegnando a ogni occorrenza di una parola una rappresentazione che dipende dal contesto in cui appare. Matematicamente, invece di una funzione $\phi: W \rightarrow \mathbb{R}^n$, abbiamo una funzione $\phi: W \times C \rightarrow \mathbb{R}^n$, dove C rappresenta lo spazio dei contesti possibili. Questa estensione permette di catturare fenomeni come la polisemia e la disambiguazione contestuale.

Nel modello BERT (Bidirectional Encoder Representations from Transformers), la rappresentazione di una parola w_i in un contesto $(w_1, \dots, w_i, \dots, w_n)$ è ottenuta attraverso una serie di trasformazioni basate su meccanismi di attenzione. Lo strato di attenzione calcola una rappresentazione pesata del contesto secondo la formula

$\text{Attention}(Q, K, V) = \text{softmax}(QK^T / \sqrt{d_k})V$, dove Q, K, V sono matrici di query, chiavi e valori derivate dalle rappresentazioni degli strati precedenti, e d_k è la dimensionalità delle chiavi. Questo meccanismo permette a ogni parola di "prestare attenzione" selettivamente ad altre parole nel contesto, modulando dinamicamente la propria rappresentazione.

La struttura geometrica degli spazi contestualizzati è significativamente più complessa di quella degli spazi statici. Per una data parola polisemica, i suoi embedding contestualizzati formano un cluster nello spazio ad alta dimensione, dove sottocluster possono corrispondere a diversi sensi della parola. La geometria di questi cluster può essere studiata attraverso tecniche di analisi multivariata come l'analisi delle componenti principali (PCA) o l'embedding uniforme di approssimazione e proiezione (UMAP), che riducono la dimensionalità preservando la struttura locale dei dati.

Varietà Differenziabili e Geometria Riemanniana nei Modelli Linguistici

Varietà Differenziabili negli Spazi Semantici

La teoria delle varietà differenziabili fornisce un framework matematico rigoroso per studiare gli spazi semantici come oggetti geometrici dotati di struttura locale euclidea ma potenzialmente complessa struttura globale. Una varietà differenziabile di dimensione n è uno spazio topologico M localmente omeomorfo a $\mathbb{R}^n$, dotato di una struttura che permette di definire in modo consistente nozioni di differenziabilità. Nel contesto degli spazi semantici, questa struttura emerge naturalmente quando consideriamo famiglie parametriche di rappresentazioni o traiettorie di ottimizzazione durante l'apprendimento.

Formalmente, una varietà differenziabile M è dotata di un atlante, ovvero una collezione di carte $(U_\alpha, \varphi_\alpha)$ dove U_α è un aperto di M e $\varphi_\alpha: U_\alpha \rightarrow \mathbb{R}^n$ è un omeomorfismo sull'immagine. Le funzioni di transizione $\varphi_\beta \circ \varphi_\alpha^{-1}: \varphi_\alpha(U_\alpha \cap U_\beta) \rightarrow \varphi_\beta(U_\alpha \cap U_\beta)$ sono richieste essere differenziabili (tipicamente di classe C^∞). Questa struttura permette di trasportare localmente lo spazio semantico in coordinate euclidee, dove le tecniche del calcolo differenziale standard diventano applicabili.

La sfera unitaria S^{n-1} , su cui vivono gli embedding normalizzati, è l'esempio prototipico di varietà differenziabile non banale negli spazi semantici. Un atlante standard per S^{n-1} può essere costruito attraverso proiezioni stereografiche dai poli nord e sud. La proiezione stereografica dal polo nord $(0, \dots, 0, 1)$ mappa un punto $(x_1, \dots, x_n) \in S^{n-1}$ in $(x_1/(1-x_n), \dots, x_{n-1}/(1-x_n)) \in \mathbb{R}^{n-1}$. La funzione di transizione tra le due proiezioni stereografiche è l'inversione $x \mapsto x/||x||^2$, una trasformazione conforme che preserva angoli ma non distanze.

Lo spazio tangente $T_p M$ in un punto p di una varietà differenziabile è lo spazio vettoriale delle derivate direzionali in p , ovvero delle velocità di curve differenziabili passanti per p . Per la sfera S^{n-1} , abbiamo già osservato che $T_p S^{n-1} = \{v \in \mathbb{R}^n : v \cdot p = 0\}$, lo spazio dei vettori ortogonali a p . La collezione di tutti gli spazi tangenti forma il fibrato tangente $TM = \bigcup_{p \in M} \{p\} \times T_p M$, una varietà differenziabile di dimensione $2n$ dotata di una proiezione naturale $\pi: TM \rightarrow M$.

Un campo vettoriale su M è una sezione del fibrato tangente, ovvero una funzione $X: M \rightarrow TM$ tale che $\pi \circ X = \text{id}_M$. In termini concreti, un campo vettoriale assegna a ogni punto $p \in M$ un vettore $X(p) \in T_p M$. Nel contesto semantico, un campo vettoriale può rappresentare una direzione di variazione sistematica del significato, come quella indotta da una specifica trasformazione semantica o da un gradiente di ottimizzazione. Il flusso di un campo vettoriale genera curve integrali che descrivono traiettorie continue attraverso lo spazio semantico.

Metriche Riemanniane e Geometria Intrinseca

Una metrica riemanniana su una varietà differenziabile M è una famiglia di prodotti scalari $g_p: T_p M \times T_p M \rightarrow \mathbb{R}$, uno per ogni punto $p \in M$, che varia in modo liscio al variare di p . La metrica riemanniana permette di misurare lunghezze di vettori tangenti, angoli tra vettori tangenti, e conseguentemente lunghezze di curve sulla varietà. Una varietà differenziabile dotata di metrica riemanniana è detta varietà riemanniana.

Per la sfera S^{n-1} immersa in $\mathbb{R}^n$, la metrica riemanniana standard è quella indotta dalla metrica euclidea ambiente. Per vettori tangenti $u, v \in T_p S^{n-1}$, il prodotto scalare riemanniano è semplicemente $g_p(u, v) = u \cdot v$, il prodotto scalare euclideo standard. Questa metrica induce una distanza intrinseca sulla sfera: la distanza tra due punti p e q è definita come l'infimo delle lunghezze delle curve che li connettono, dove la lunghezza di una curva $\gamma: [a, b] \rightarrow M$ è data da $L(\gamma) = \int_a^b \sqrt{g(\gamma'(t))(\gamma'(t))} dt$.

Le curve che localmente minimizzano la distanza sono dette geodetiche e soddisfano l'equazione delle geodetiche, un sistema di equazioni differenziali ordinarie del secondo ordine. Per la sfera, le geodetiche sono precisamente gli archi di cerchi massimi, ovvero le intersezioni della sfera con piani passanti per l'origine. Una geodetica è univocamente determinata da un punto iniziale p e una velocità iniziale $v \in T_p M$, e può essere vista come la generalizzazione del concetto di "linea retta" in geometria non euclidea.

Il tensore di curvatura di Riemann è l'oggetto fondamentale che quantifica la deviazione della geometria della varietà dalla piattezza euclidea. Per una varietà riemanniana (M, g) , il tensore di Riemann R è un campo tensoriale di tipo $(1,3)$ che misura la non commutatività della derivata covariante. In coordinate locali, le componenti del tensore sono date da $R^i_{jkl} = \partial_k \Gamma^i_{jl} - \partial_l \Gamma^i_{jk} + \Gamma^i_{mk} \Gamma^m_{jl} - \Gamma^i_{ml} \Gamma^m_{jk}$, dove Γ^i_{jk} sono i simboli di Christoffel che codificano la connessione di Levi-Civita.

La curvatura sezionale $K(\pi)$ di un 2-piano $\pi \subset T_p M$ misura la curvatura gaussiana della superficie ottenuta trasportando parallelamente π lungo geodetiche. Per la sfera S^{n-1} di raggio r , la curvatura sezionale è costante e uguale a $K = 1/r^2$. Per la sfera unitaria, $K = 1$. Questa curvatura positiva costante implica proprietà geometriche caratteristiche: triangoli geodetici hanno somma degli angoli maggiore di π , circonferenze hanno perimetro minore di $2\pi r$, e geodetiche inizialmente parallele convergono.

Il tensore di Ricci Ric è una contrazione del tensore di Riemann: $\text{Ric}_{ij} = R^k_{ikj}$. Per la sfera n -dimensionale, $\text{Ric} = (n-1)g$, dove g è il tensore metrico. La curvatura scalare, definita come la traccia del tensore di Ricci, è $S = n(n-1)$ per la sfera unitaria n -dimensionale. Queste quantità giocano un ruolo cruciale nell'equazione di Einstein della relatività generale e nella geometria globale attraverso teoremi come quello di Bonnet-Myers, che afferma che una

varietà riemanniana completa con curvatura di Ricci uniformemente limitata dal basso da una costante positiva deve essere compatta con diametro finito.

Trasporto Parallelo e Olonomia

Il trasporto parallelo è un meccanismo per confrontare vettori tangenti in punti diversi di una varietà riemanniana, essenziale per definire derivate direzionali di campi vettoriali. Data una curva $\gamma: [0,1] \rightarrow M$ e un vettore $v_0 \in T_{\gamma(0)} M$, il trasporto parallelo di v_0 lungo γ produce un campo vettoriale $V(t)$ lungo γ tale che $V(0) = v_0$ e la derivata covariante $DV/dt = 0$. L'equazione del trasporto parallelo in coordinate locali è $dV^i/dt + \Gamma^i_{jk} (d\gamma^j/dt) V^k = 0$.

Su una sfera, il trasporto parallelo lungo una geodetica mantiene costante l'angolo tra il vettore trasportato e la geodetica stessa. Tuttavia, il trasporto lungo curve chiuse non geodetiche può portare a una rotazione del vettore, un fenomeno noto come olonomia. Il gruppo di olonomia $Hol_p(M)$ in un punto p è il gruppo di tutte le trasformazioni lineari di $T_p M$ ottenibili trasportando parallelamente vettori lungo curve chiuse basate in p . Per la sfera, il gruppo di olonomia è $SO(n)$, il gruppo delle rotazioni ortogonali.

L'olonomia ha una profonda connessione con la curvatura attraverso il teorema di Ambrose-Singer, che afferma che l'algebra di Lie del gruppo di olonomia è generata dalle curvature lungo tutte le possibili geodetiche. Nel contesto semantico, l'olonomia può essere interpretata come la dipendenza dal percorso delle trasformazioni semantiche: muoversi attraverso lo spazio semantico lungo percorsi diversi può produrre trasformazioni diverse degli attributi semantici, anche quando i punti di partenza e arrivo coincidono.

Geometria Iperbolica e Gerarchie Lessicali

Mentre la sfera fornisce un modello naturale per gli spazi semantici in dimensione finita, la geometria iperbolica offre un framework alternativo particolarmente adatto per rappresentare strutture gerarchiche. Lo spazio iperbolico di dimensione n , denotato H^n , è una varietà riemanniana di curvatura sezionale costante negativa, che possiede proprietà geometriche radicalmente diverse da quelle dello spazio euclideo e della sfera.

Esistono diversi modelli equivalenti dello spazio iperbolico, ciascuno con i propri vantaggi computazionali. Il modello di Poincaré del disco rappresenta H^n come il disco unitario aperto $D^n = \{x \in \mathbb{R}^n : \|x\| < 1\}$ dotato della metrica riemanniana $g_x = (4/(1-\|x\|^2)^2) g_E$, dove g_E è la metrica euclidea standard. In questo modello, le geodetiche sono archi di cerchi ortogonali al bordo del disco (o diametri). La distanza iperbolica tra due punti $x, y \in D^n$ è data da $d_H(x, y) = \operatorname{arccosh}(1 + 2\|x-y\|^2/((1-\|x\|^2)(1-\|y\|^2)))$.

Il modello di Poincaré del semispazio superiore rappresenta H^n come $H^n = \{(x_1, \dots, x_n) \in \mathbb{R}^n : x_n > 0\}$ con metrica $g = (1/x_n^2) g_E$. Le geodetiche in questo modello sono semicerchi e semirette verticali ortogonali all'iperpiano $x_n = 0$. Il modello del lorentziano rappresenta H^n come un iperboloide nello spazio di Minkowski $\mathbb{R}^{n,1}$, l'insieme $\{x \in \mathbb{R}^{n+1} : -x_0^2 + x_1^2 + \dots + x_n^2 = -1, x_0 > 0\}$ dotato della metrica indotta dalla forma quadratica di Minkowski.

La caratteristica distintiva della geometria iperbolica è la crescita esponenziale del volume con il raggio. Mentre in spazio euclideo il volume di una palla di raggio r cresce come r^n , e su una sfera rimane limitato, in spazio iperbolico il volume cresce come $e^{\{(n-1)r\}}$. Questa crescita esponenziale rende lo spazio iperbolico ideale per rappresentare strutture ad albero e gerarchie, dove il numero di nodi cresce esponenzialmente con la profondità.

Gli embedding iperbolici per rappresentazioni lessicali sono stati introdotti per catturare relazioni gerarchiche come iperonimia e meronimia. In un embedding iperbolico, parole che sono iperonimi (concetti più generali) sono posizionate vicino al centro del modello di Poincaré, mentre iponimi (concetti più specifici) sono posizionati più vicino al bordo. La distanza dal centro può essere interpretata come una misura di specificità concettuale. La struttura gerarchica emerge dalla geometria: i discendenti di un concetto formano un settore angolare nello spazio, e la distanza iperbolica cresce logicamente con la profondità gerarchica.

Formalmente, dato un albero con fattore di ramificazione b e profondità d , è possibile embedderlo isometricamente (preservando le distanze) nello spazio iperbolico di dimensione 2 con distorsione limitata, mentre in spazio euclideo la dimensione richiesta cresce almeno linearmente con il numero di nodi. Questo risultato, basato su lavori di Gromov e altri, fornisce una giustificazione teorica per l'uso della geometria iperbolica nella rappresentazione di gerarchie lessicali.

L'ottimizzazione riemanniana in spazio iperbolico richiede tecniche specifiche che rispettino la geometria della varietà. Il gradiente riemanniano di una funzione $f: H^n \rightarrow \mathbb{R}$ è definito come l'unico vettore tangente $\text{grad}^R f$ tale che $g(\text{grad}^R f, v) = df(v)$ per ogni vettore tangente v . Nel modello di Poincaré, il gradiente riemanniano è legato al gradiente euclideo dalla relazione $\text{grad}^R f = ((1 - \|x\|^2)^2 / 4) \text{grad}^E f$. L'aggiornamento dei parametri segue geodetiche piuttosto che linee rette, attraverso la mappa esponenziale $\exp_x: T_x M \rightarrow M$ che mappa vettori tangenti in x a punti della varietà.

Operazioni Geometriche negli Spazi Iperbolici

Le operazioni fondamentali negli spazi iperbolici richiedono definizioni che rispettino la geometria non euclidea. L'addizione di Möbius nel modello di Poincaré del disco è definita come $x \oplus y = ((1 + 2\langle x, y \rangle + \|y\|^2)x + (1 - \|x\|^2)y) / (1 + 2\langle x, y \rangle + \|x\|^2\|y\|^2)$, dove $\langle \cdot, \cdot \rangle$ denota il prodotto scalare euclideo. Questa operazione non è commutativa, riflettendo l'assenza di traslazioni isometriche nello spazio iperbolico.

La moltiplicazione scalare iperbolica è definita attraverso la mappa esponenziale e logaritmica. Data uno scalare $r \in \mathbb{R}$ e un punto $x \in D^n$, la moltiplicazione $r \otimes x$ è definita come $\exp_0(r \log_0(x))$, dove $\log_0(x) = (\text{arctanh}(\|x\|) / \|x\|) x$ è la mappa logaritmica che inverte l'esponenziale. Per vettori di norma unitaria nella metrica euclidea, questa operazione si riduce a $r \otimes x = \tanh(r \text{arctanh}(\|x\|)) (x / \|x\|)$.

Il centroide o baricentro di un insieme di punti $\{x_1, \dots, x_n\}$ in spazio iperbolico può essere definito come il punto che minimizza la somma delle distanze iperboliche al quadrato: $\text{argmin}_x \sum_i d_H(x, x_i)^2$. A differenza dello spazio euclideo, dove il baricentro ha un'espressione esplicita come media aritmetica, in spazio iperbolico deve essere calcolato iterativamente attraverso algoritmi di punto fisso o discesa del gradiente riemanniano.

Applicazioni alla Composizione Semantica

La geometria iperbolica fornisce un framework naturale per la composizione semantica gerarchica. Consideriamo la composizione di un sostantivo con un aggettivo, come in "cane grande". Se "cane" è rappresentato da un punto c nello spazio iperbolico e "grande" modifica il livello di specificità spostando il punto verso il bordo (aumentando la specificità), la composizione può essere realizzata attraverso una traslazione iperbolica nella direzione appropriata.

Formalmente, la composizione può essere definita come $\phi(\text{aggettivo}, \text{sostantivo}) = \exp_{\text{sostantivo}}(f(\text{aggettivo}))$, dove f è una funzione che mappa l'aggettivo in un vettore nello spazio tangente al sostantivo, e $\exp$ è la mappa esponenziale. La funzione f può essere appresa attraverso dati di training che associano coppie (aggettivo, sostantivo) alle loro composizioni attestate. Questo approccio permette di catturare sia modificazioni che preservano il livello gerarchico (spostamento orizzontale) sia modificazioni che cambiano la specificità (spostamento radiale).

Per composizioni più complesse, come frasi intere, si possono utilizzare reti neurali che operano direttamente nello spazio iperbolico. Le operazioni di aggregazione, come la media o la somma pesata, devono essere sostituite dalle loro controparti iperboliche. Un modello ricorrente iperbolico aggiorna il suo stato attraverso $h_t = \exp_{\{h_{t-1}\}}(W \log_{\{h_{t-1}\}}(x_t))$, dove x_t è l'input al tempo t , h_{t-1} è lo stato precedente, e W è una matrice di pesi da apprendere.

Geometria del Prodotto e Spazi Misti

Per catturare simultaneamente strutture gerarchiche e simmetriche, sono stati proposti spazi prodotto che combinano geometrie diverse. Uno spazio prodotto $M_1 \times M_2 \times \dots \times M_k$, dove ciascun M_i è una varietà riemanniana (euclidea, sferica, o iperbolica), eredita una metrica prodotto definita come $d((x_1, \dots, x_k), (y_1, \dots, y_k))^2 = \sum_i d_i(x_i, y_i)^2$, dove d_i è la metrica su M_i .

Un embedding in uno spazio prodotto assegna a ogni parola una tupla di vettori, ciascuno vivente in una diversa geometria. Questo permette di separare diversi aspetti del significato: componenti sferiche possono catturare attributi simmetrici come il genere o la polarità emotiva, componenti iperboliche possono catturare relazioni gerarchiche come iperonimia, e componenti euclidee possono rappresentare aspetti continui non strutturati.

L'ottimizzazione in spazi prodotto richiede il calcolo di gradienti rispetto a ciascuna componente geometrica. Il gradiente riemanniano totale è la somma dei gradienti nelle singole componenti, ciascuno calcolato secondo la geometria corrispondente. La retropropagazione deve tenere conto delle diverse strutture metriche, utilizzando le mappe esponenziali e logaritmiche appropriate per ciascuna geometria.

Divergenze Statistiche come Distanze Riemanniane

Le divergenze tra distribuzioni di probabilità possono essere interpretate geometricamente come distanze su varietà statistiche. La geometria dell'informazione, sviluppata da Amari e

altri, studia le famiglie di distribuzioni di probabilità come varietà differenziabili dotate di metriche naturali derivate dalla teoria dell'informazione.

La metrica di Fisher è la metrica riemanniana fondamentale sulla varietà statistica. Per una famiglia parametrica di distribuzioni $p(x|\theta)$, dove $\theta = (\theta_1, \dots, \theta_n)$ sono i parametri, il tensore metrico di Fisher è definito come $g_{\{ij\}}(\theta) = E[\partial \log p(x|\theta)/\partial \theta_i \cdot \partial \log p(x|\theta)/\partial \theta_j] = -E[\partial^2 \log p(x|\theta)/\partial \theta_i \partial \theta_j]$, dove l'aspettazione è rispetto a $p(x|\theta)$. Questa metrica è invariante rispetto a riparametrizzazioni sufficienti, una proprietà desiderabile per una geometria intrinseca.

La divergenza di Kullback-Leibler può essere approssimata localmente dalla distanza di Fisher. Per distribuzioni infinitesimalmente vicine, $D_{KL}(p(\cdot|\theta) || p(\cdot|\theta+d\theta)) \approx (1/2) \sum_{\{ij\}} g_{\{ij\}}(\theta) d\theta_i d\theta_j$. Questa relazione mostra che la divergenza KL induce una struttura geometrica sulla varietà delle distribuzioni, dove la metrica di Fisher gioca il ruolo di tensore metrico.

Le connessioni affini sulla varietà statistica, in particolare la connessione esponenziale e la connessione mistura, generalizzano la connessione di Levi-Civita e permettono di definire geodetiche che minimizzano diverse divergenze. La geometria duale di Amari introduce due strutture affini coniugate, che portano a una ricca struttura geometrica con applicazioni all'inferenza statistica e all'apprendimento automatico.

Analisi Tensoriale della Composizione Semantica

Fondamenti della Rappresentazione Tensoriale

La composizione semantica, ovvero il processo attraverso cui i significati di unità linguistiche elementari si combinano per formare significati complessi, richiede strutture matematiche più ricche dei semplici vettori. I tensori, generalizzazioni multidimensionali dei vettori e delle matrici, forniscono un framework naturale per rappresentare interazioni complesse tra elementi semantici. Un tensore di ordine k è un elemento di uno spazio tensoriale $V_1 \otimes V_2 \otimes \dots \otimes V_k$, dove ciascun V_i è uno spazio vettoriale.

Formalmente, un tensore T di tipo (r,s) su uno spazio vettoriale V è un'applicazione multilineare $T: V^* \times \dots \times V^* \times V \times \dots \times V \rightarrow \mathbb{R}$, dove V^* è lo spazio duale di V , e ci sono r copie di V^* e s copie di V . I vettori sono tensori di tipo $(1,0)$, i covettori (funzionali lineari) sono tensori di tipo $(0,1)$, e le matrici possono essere viste come tensori di tipo $(1,1)$ o $(2,0)$ a seconda dell'interpretazione. In coordinate, un tensore di tipo (r,s) è rappresentato da un array con r indici controvarianti e s indici covarianti: $T^{\{i_1 \dots i_r\}}_{\{j_1 \dots j_s\}}$.

Nel contesto semantico, un tensore di ordine tre può rappresentare la composizione di tre parole. Per esempio, il tensore $T \in \mathbb{R}^n \otimes \mathbb{R}^n \otimes \mathbb{R}^n$ può codificare relazioni ternarie tra parole, dove $T_{\{ijk\}}$ quantifica l'intensità dell'associazione tra le parole i , j , e k . Più concretamente, se vogliamo comporre un aggettivo, un sostantivo e un verbo per formare il significato di una frase semplice, possiamo utilizzare un tensore di composizione che mappa la tripla di vettori di parole in un vettore di frase.

Il prodotto tensoriale di due spazi vettoriali V e W , denotato $V \otimes W$, è lo spazio vettoriale generato da simboli formali $v \otimes w$ per ogni $v \in V$ e $w \in W$, soggetti alle relazioni di bilinearità: $(\alpha v_1 + \beta v_2) \otimes w = \alpha(v_1 \otimes w) + \beta(v_2 \otimes w)$ e $v \otimes (\alpha w_1 + \beta w_2) = \alpha(v \otimes w_1) + \beta(v \otimes w_2)$. La dimensione di $V \otimes W$ è il prodotto delle dimensioni di V e W . Se $\{e_i\}$ è una base per V e $\{f_j\}$ è una base per W , allora $\{e_i \otimes f_j\}$ forma una base per $V \otimes W$.

Modelli Tensoriali per la Composizione Lessicale

Il modello tensoriale più semplice per la composizione di due parole è quello additivo, dove il significato della composizione è la somma (o media) dei vettori delle parole componenti: $\phi(w_1, w_2) = (v_1 + v_2)/2$. Questo modello, sebbene estremamente semplice, cattura l'intuizione che il significato composto dovrebbe preservare aspetti di entrambi i significati componenti. Tuttavia, è chiaramente inadeguato per catturare la non commutatività e la non associatività della composizione linguistica: "uomo mangia cane" ha un significato diverso da "cane mangia uomo".

Il modello moltiplicativo utilizza il prodotto componente per componente (Hadamard): $\phi(w_1, w_2) = v_1 \odot v_2$, dove $(v_1 \odot v_2)_i = v_{1,i} \cdot v_{2,i}$. Questo modello permette interazioni tra le dimensioni dei vettori, ma rimane commutativo e non introduce nuove direzioni nello spazio semantico. Una variante pesata $\phi(w_1, w_2) = \alpha(v_1 \odot v_2) + \beta v_1 + \gamma v_2$ combina componenti moltiplicative e additive, dove i pesi α, β, γ possono essere appresi.

Il modello basato su matrici di composizione assegna a ciascuna parola non solo un vettore v ma anche una matrice M che cattura come quella parola modifica il significato di altre parole. La composizione di due parole è data da $\phi(w_1, w_2) = M_1 v_2$ o $\phi(w_1, w_2) = v_1^T M_2 v_2$ o ancora $\phi(w_1, w_2) = f(M_1 v_2, M_2 v_1)$, dove f è una funzione di combinazione, tipicamente additiva o moltiplicativa. Questo modello permette asimmetria: aggettivi possono avere matrici che modificano sostantivi in modo diverso da come i sostantivi modificano aggettivi.

Il modello di tensore completo rappresenta la composizione attraverso un tensore di ordine tre $T \in \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^n$, dove la composizione di parole con vettori v_1 e v_2 è data dalla contrazione $\phi(w_1, w_2) = T \times_i v_1 \times_j v_2$, dove $\times_i$ denota la contrazione lungo il modo i . In notazione tensoriale, questo si scrive come $(\phi(w_1, w_2))_k = \sum_{ij} T_{ijk} v_{1,i} v_{2,j}$. Il tensore T ha n^3 parametri, il che lo rende computazionalmente proibitivo per dimensioni elevate.

Per ridurre il numero di parametri, si utilizzano decomposizioni tensoriali. La decomposizione di Tucker rappresenta un tensore come $T = G \times_1 U_1 \times_2 U_2 \times_3 U_3$, dove G è un tensore nucleo di dimensione ridotta e le U_i sono matrici fattore. Questo riduce i parametri da n^3 a $r^3 + 3nr$, dove $r < n$ è il rango della decomposizione. La decomposizione CANDECOMP/PARAFAC (CP) rappresenta T come somma di r tensori di rango uno: $T = \sum_{k=1}^r a_k \otimes b_k \otimes c_k$, dove a_k, b_k, c_k sono vettori. Questo riduce ulteriormente i parametri a $3nr$.

Composizione Sintattico-Semantica Ricorsiva

Per composizioni gerarchiche che seguono strutture sintattiche ad albero, i modelli ricorsivi applicano ripetutamente operazioni di composizione binaria seguendo la struttura dell'albero sintattico. Dato un albero sintattico con parole $w_1, \dots, w_n$ alle foglie e nodi interni che

rappresentano costituenti sintattici, il significato viene calcolato bottom-up attraverso $\phi(\text{tree}) = f(\phi(\text{left_subtree}), \phi(\text{right_subtree}))$, dove f è una funzione di composizione.

Le Recursive Neural Networks (RNN) implementano questo schema utilizzando una rete neurale per la funzione di composizione. Dato un nodo con figli sinistro e destro con rappresentazioni v_L e v_R , la rappresentazione del nodo genitore è $v_{\text{parent}} = \tanh(W[v_L; v_R] + b)$, dove W è una matrice di pesi, b è un bias, e $[v_L; v_R]$ denota la concatenazione dei vettori. I pesi W e b sono condivisi attraverso tutti i nodi dell'albero, permettendo la generalizzazione a alberi di forma arbitraria.

Le Recursive Neural Tensor Networks (RNTN) introducono interazioni bilineari tra i vettori figli attraverso un tensore: $v_{\text{parent}} = \tanh(v_L^T T^{\{1:n\}} v_R + W[v_L; v_R] + b)$, dove $T^{\{1:n\}} \in R^{d \times d \times n}$ è un tensore tridimensionale e l'operazione $v_L^T T^{\{i\}} v_R$ produce lo scalare $\sum_{j,k} v_{L,j} T^{\{i\}}_{j,k} v_{R,k}$. Questa interazione tensoriale permette al modello di catturare relazioni complesse tra i costituenti, come negazioni o modificazioni intensive.

La composizione dipendente dalla categoria sintattica può essere realizzata assegnando matrici di composizione diverse a diverse categorie sintattiche. Un aggettivo ha una matrice M_{adj} , un sostantivo ha M_{noun} , un verbo M_{verb} , e così via. La composizione $\text{adj} + \text{noun}$ utilizza M_{adj} per modificare v_{noun} , mentre $\text{verb} + \text{noun}$ utilizza M_{verb} . Questo approccio si ispira alla grammatica categoriale, dove ogni parola è associata a un tipo sintattico che determina come può combinarsi con altre parole.

Algebra Tensoriale e Calcolo Covariante

Il calcolo tensoriale fornisce gli strumenti per manipolare tensori su varietà, generalizzando il calcolo vettoriale ordinario. La derivata covariante ∇ è un'operazione che permette di differenziare campi tensoriali rispettando la geometria della varietà. Per un campo vettoriale V e una direzione u , la derivata covariante $\nabla_u V$ è un vettore che misura come V varia nella direzione u tenendo conto della curvatura della varietà.

In coordinate locali, la derivata covariante di un campo vettoriale V^i è data da $(\nabla_j V)^i = \partial V^i / \partial x^j + \Gamma^i_{jk} V^k$, dove Γ^i_{jk} sono i simboli di Christoffel della connessione. Per un covettore ω_i , la derivata covariante è $(\nabla_j \omega)_i = \partial \omega_i / \partial x^j - \Gamma^k_{ji} \omega_k$. Queste formule si generalizzano a tensori di ordine arbitrario attraverso la regola di Leibniz tensoriale.

Nel contesto semantico, la derivata covariante permette di definire come un significato varia quando ci si muove nello spazio semantico lungo una direzione specificata. Se $\phi(t)$ è una curva nello spazio semantico e V è un campo vettoriale che rappresenta un attributo semantico, la derivata covariante DV/dt lungo ϕ misura come questo attributo cambia, corretto per la geometria dello spazio.

Il parallelogramma di Lie-trasporto fornisce un metodo per trasportare vettori tra punti non infinitesimamente vicini. Date due curve $\gamma_1(t)$ e $\gamma_2(s)$ che si intersecano in $p = \gamma_1(0) = \gamma_2(0)$, e dato un vettore v in p , il Lie-trasporto di v lungo prima γ_1 fino a $q = \gamma_1(1)$ e poi indietro lungo γ_2 fino a p può differire da v di una quantità che dipende dalla curvatura. Questa discrepanza è misurata dalla curvatura di Riemann e ha interpretazioni semantiche in termini di dipendenza dal percorso delle trasformazioni di significato.

Decomposizioni Tensoriali e Spazi Latenti

La decomposizione tensoriale permette di scoprire strutture latenti in dati semantici multidimensionali. Consideriamo un tensore di co-occorrenza tripla $T \in \mathbb{R}^{n \times n \times n}$, dove $T_{\{ijk\}}$ quantifica la frequenza con cui le parole i, j, k co-occorrono in qualche configurazione contestuale. La decomposizione CP di questo tensore $T \approx \sum_{r=1}^R a_r \otimes b_r \otimes c_r$ rivela R fattori latenti, ciascuno rappresentato da una tripla di vettori (a_r, b_r, c_r) .

Questi fattori latenti possono essere interpretati come "proto-concetti" o dimensioni semantiche fondamentali che sottendono le co-occorrenze osservate. Se le parole i, j, k co-occorrono frequentemente, questo è spiegato dal fatto che esse hanno alto carico su fattori latenti comuni. La decomposizione CP è unica sotto condizioni generali, a differenza della decomposizione matriciale SVD, il che la rende particolarmente attraente per l'interpretabilità.

La decomposizione di Tucker $T \approx G \times_1 U \times_2 V \times_3 W$ rappresenta il tensore attraverso un tensore nucleo G e tre matrici fattore. Mentre la decomposizione CP assume che i fattori siano identici nei tre modi, Tucker permette fattori diversi, catturando così interazioni più complesse. Il tensore nucleo G codifica le interazioni tra i fattori latenti nei diversi modi.

Gli algoritmi per il calcolo di queste decomposizioni includono metodi iterativi come Alternating Least Squares (ALS), che fissa tutti i fattori tranne uno e risolve un problema di minimi quadrati per quel fattore, ciclando attraverso tutti i fattori fino alla convergenza. Altri approcci includono metodi basati su gradienti, decomposizioni gerarchiche, e tecniche tensoriali randomizzate che sfruttano proiezioni casuali per ridurre la complessità computazionale.

Campi Tensoriali Semantici

Un campo tensoriale su uno spazio semantico M è una sezione del fibrato tensoriale, ovvero un'assegnazione che associa a ogni punto $p \in M$ un tensore T_p nel corrispondente spazio tensoriale in p . Nel contesto semantico, un campo tensoriale di ordine due (matrice) può rappresentare come le relazioni tra attributi semantici variano attraverso lo spazio lessicale.

Per esempio, consideriamo un campo matriciale $A(p)$ che descrive come un certo attributo semantico (come "intensità emotiva") si trasforma linearmente in funzione della posizione p nello spazio semantico. Se v è un vettore che rappresenta una particolare configurazione di attributi, $A(p)v$ descrive come questa configurazione viene trasformata nel punto p . La divergenza del campo matriciale, $\nabla \cdot A = \sum_i \nabla_i A^i_j$, misura la "sorgente" locale della trasformazione.

Il tensore di deformazione ϵ di un campo vettoriale V su una varietà riemanniana è definito come $\epsilon_{ij} = (1/2)(\nabla_i V_j + \nabla_j V_i)$, dove ∇ è la derivata covariante e gli indici sono abbassati usando il tensore metrico. Il tensore di deformazione è simmetrico e rappresenta la parte simmetrica del gradiente di V . Nel contesto semantico, se V rappresenta un "flusso" di significato, ϵ descrive come questo flusso deforma localmente lo spazio.

Il tensore di rotazione o vorticità $\Omega = (1/2)(\nabla_i V_j - \nabla_j V_i)$ cattura la parte antisimmetrica del gradiente, descrivendo rotazioni locali del flusso. In tre dimensioni, questo è equivalente

al rotore $\nabla \times V$. Un campo vettoriale con vorticità nulla è detto irrotazionale e può essere espresso come gradiente di una funzione potenziale.

Tensori di Riemann e Invarianti Semantici

Il tensore di curvatura di Riemann R , già introdotto, è il tensore fondamentale che codifica tutta l'informazione sulla curvatura intrinseca della varietà. Da esso possono essere derivati vari tensori e scalari che forniscono informazioni geometriche globali e locali. Il tensore di Ricci $\text{Ric}_{\{ij\}} = R^k_{\{ik\}}$ è una contrazione del tensore di Riemann e appare nell'equazione di Einstein. La curvatura scalare $S = g^{\{ij\}} \text{Ric}_{\{ij\}}$ è una ulteriore contrazione che fornisce un singolo numero che descrive la curvatura in ogni punto.

Il tensore di Weyl W è la parte del tensore di Riemann che è traceless, ovvero $W^i_{\{jkl\}} = R^i_{\{jkl\}} - (1/(n-2))(\delta^i_k \text{Ric}_{\{jl\}} - \delta^i_l \text{Ric}_{\{jk\}} + g_{\{jl\}} \text{Ric}^i_k - g_{\{jk\}} \text{Ric}^i_l) + S/((n-1)(n-2))(\delta^i_k g_{\{jl\}} - \delta^i_l g_{\{jk\}})$. Il tensore di Weyl vanisce identicamente in dimensione 2 e 3, ma in dimensioni superiori rappresenta la curvatura "genuinamente multidimensionale" che non può essere ridotta a curvature di superfici bidimensionali.

Nel contesto semantico, questi invarianti curvaturali possono essere interpretati come misure intrinseche della complessità e della struttura dello spazio dei significati. Regioni con alta curvatura scalare positiva corrispondono a campi semantici densamente interconnessi, dove i significati sono fortemente correlati. Regioni con curvatura negativa possono corrispondere a "selle semantiche", ovvero punti dove significati divergenti si incontrano.

Topological Data Analysis Applicata agli Spazi Semantici

La Topological Data Analysis (TDA) è un campo che studia la forma dei dati attraverso strumenti della topologia algebrica, in particolare l'omologia persistente. L'idea fondamentale è di costruire una famiglia di spazi topologici indicizzata da un parametro di scala, e studiare come le caratteristiche topologiche (componenti connesse, cicli, cavità) persistono attraverso le scale.

Dato un insieme di punti X nello spazio semantico, il complesso di Vietoris-Rips $R_\epsilon(X)$ al parametro ϵ è il complesso simpliciale i cui k -simplessi sono formati da $(k+1)$ -uple di punti in X tali che ogni coppia ha distanza al massimo ϵ . Per ϵ piccolo, $R_\epsilon(X)$ consiste di punti isolati. Aumentando ϵ , i punti iniziano a connettersi attraverso spigoli (1-simplessi), formano triangoli (2-simplessi), tetraedri (3-simplessi), e così via.

L'omologia persistente traccia l'evoluzione dei gruppi di omologia $H_k(R_\epsilon(X))$ al variare di ϵ . Il k -esimo gruppo di omologia conta essenzialmente i k -cicli indipendenti: H_0 conta le componenti connesse, H_1 conta i loop indipendenti (cicli che non sono bordi di superfici), H_2 conta le cavità tridimensionali, e così via. Un generatore di H_k "nasce" a un certo parametro ϵ_{birth} quando appare per la prima volta, e "muore" a ϵ_{death} quando viene riempito o connesso ad altri generatori.

Il persistence diagram è una rappresentazione delle caratteristiche topologiche come punti nel piano $(\epsilon_{\text{birth}}, \epsilon_{\text{death}})$. Punti lontani dalla diagonale corrispondono a caratteristiche persistenti, che sono considerate "significative", mentre punti vicini alla diagonale sono

attribuiti a rumore. Il barcode è una rappresentazione equivalente dove ogni caratteristica è un intervallo $[\varepsilon_{\text{birth}}, \varepsilon_{\text{death}})$.

Applicata agli spazi semantici, l'omologia persistente può rivelare la struttura modulare dei campi semantici. Componenti connesse in H_0 corrispondono a cluster di parole semanticamente correlate. I generatori di H_1 , ovvero loop persistenti, possono corrispondere a relazioni cicliche tra concetti, come sinonimie circolari o catene di associazioni che ritornano su se stesse. Cavità in H_2 possono indicare strutture semantiche "cave", ovvero regioni dello spazio circondate da concetti ma prive di concetti centrali.

Mapper e Visualizzazione Topologica

L'algoritmo Mapper, introdotto da Singh, Mémoli e Carlsson, fornisce un metodo per costruire approssimazioni simpliciali della topologia di dati ad alta dimensione. L'idea è di proiettare i dati su uno spazio di dimensione inferiore attraverso una funzione filtro $f: X \rightarrow \mathbb{R}^m$, ricoprire l'immagine con insiemi aperti, e costruire un grafo le cui vertici corrispondono a cluster nelle preimmagini dei set aperti, con archi che connettono cluster con intersezione non vuota.

Formalmente, dato un insieme di dati X , una funzione filtro $f: X \rightarrow \mathbb{R}^m$, un ricoprimento aperto $U = \{U_\alpha\}$ di $f(X)$, e un algoritmo di clustering, Mapper procede come segue. Per ogni U_α in U , si considera la preimmagine $f^{-1}(U_\alpha) \subset X$. Si applica l'algoritmo di clustering a $f^{-1}(U_\alpha)$, ottenendo cluster $C_{\alpha,1}, \dots, C_{\alpha,k_\alpha}$. Ogni cluster diventa un nodo nel grafo Mapper. Due nodi $C_{\alpha,i}$ e $C_{\beta,j}$ sono connessi da un arco se $C_{\alpha,i} \cap C_{\beta,j} \neq \emptyset$.

La scelta della funzione filtro è cruciale e dipende dall'applicazione. Per spazi semantici, funzioni naturali includono la distanza da un punto di riferimento (per esplorare la struttura radiale), la densità locale (per identificare cluster densi), le prime componenti principali (per catturare direzioni di massima varianza), o funzioni legate a attributi semantici specifici come polarità, concretezza, o frequenza.

Il grafo Mapper fornisce una rappresentazione semplificata della topologia dei dati che preserva caratteristiche essenziali come componenti connesse, loop, e ramificazioni. Può essere visualizzato e esplorato interattivamente, con nodi colorati secondo statistiche aggregate (come la media di un attributo nei punti del cluster). Questa visualizzazione può rivelare strutture gerarchiche, transizioni graduali tra campi semantici, e punti di biforcazione dove un campo semantico si divide in sottocampi.

Persistence Landscapes e Metodi Statistici

Per confrontare diagrammi di persistenza e condurre analisi statistica su di essi, sono necessarie rappresentazioni funzionali o vettoriali. Il persistence landscape trasforma un diagramma di persistenza in una funzione $\lambda: \mathbb{N} \times \mathbb{R} \rightarrow \mathbb{R}$ definita come $\lambda_k(t) = k$ -esimo valore più grande di $\min(t - \text{birth}_i, \text{death}_i - t)^+$, dove $(x)^+ = \max(x, 0)$ e l'insieme $\{(\text{birth}_i, \text{death}_i)\}$ sono i punti nel diagramma.

Il persistence landscape è una funzione continua, lineare a tratti, che codifica l'intero diagramma di persistenza. Può essere discretizzato per formare un vettore di dimensione

finita sommando i valori su una griglia. Lo spazio dei persistence landscapes è uno spazio di Banach dove è possibile definire medie, varianze, e condurre test statistici. Questo permette di rispondere a domande come: "Gli spazi semantici di lingue diverse hanno topologie significativamente diverse?" o "Come cambia la topologia dello spazio semantico durante l'addestramento di un modello?".

Il persistence image è un'altra rappresentazione vettoriale che trasforma un diagramma di persistenza in un'immagine bidimensionale. Il piano (birth, persistence) viene discretizzato in una griglia di pixel, e a ogni punto $(b_i, d_i - b_i)$ nel diagramma è associata una distribuzione gaussiana bidimensionale. Sommando questi contributi gaussiani su tutti i punti e pesando per la persistenza, si ottiene un'immagine che può essere usata come input per algoritmi di machine learning standard.

Coomologia e Strutture Duali

Mentre l'omologia studia cicli (catene che sono chiuse ma non bordi), la coomologia studia cocicli (funzionali lineari su catene che annullano i bordi). I gruppi di coomologia H^k sono duali ai gruppi di omologia H_k attraverso il teorema dei coefficienti universali. La coomologia ha una struttura algebrica più ricca dell'omologia: il prodotto cup $\cup$ fornisce ai gruppi di coomologia una struttura di anello graduato.

Nel contesto semantico, un cociclo può essere interpretato come un "osservabile topologico", ovvero una quantità che può essere misurata su configurazioni semantiche e che è invariante sotto deformazioni continue. Per esempio, un 1-cociclo assegna numeri agli archi di un complesso in modo consistente intorno a loop, e può rappresentare un flusso conservativo di informazione semantica.

Il prodotto cup permette di costruire cocicli di ordine superiore combinando cocicli di ordine inferiore. Se $\alpha \in H^p$ e $\beta \in H^q$, allora $\alpha \cup \beta \in H^{p+q}$. Questa operazione cattura interferenze e interazioni tra diversi osservabili topologici. Nel contesto semantico, il prodotto cup potrebbe modellare come diverse dimensioni di significato interagiscono per produrre fenomeni semantici complessi.

Modelli Geometrici della Polisemia e Fenomeni Semantici Complessi

La Polisemia come Fenomeno Geometrico

La polisemia, ovvero l'esistenza di molteplici significati correlati per una singola forma lessicale, rappresenta una delle sfide più profonde per i modelli spaziali del significato. Mentre gli embedding statici classici assegnano un unico punto nello spazio a ogni forma lessicale, catturando una sorta di significato "medio" attraverso tutti i suoi usi, la realtà della polisemia richiede rappresentazioni più articolate che possano distinguere e mettere in relazione i diversi sensi di una parola.

Da un punto di vista geometrico, la polisemia può essere concettualizzata in diversi modi. Un primo approccio consiste nel rappresentare una parola polisemica non come un singolo

punto, ma come una regione o una sottovarietà dello spazio semantico. I diversi sensi della parola corrispondono a diversi punti o regioni all'interno di questa sottovarietà, e la struttura geometrica della sottovarietà codifica le relazioni tra i sensi. Per esempio, una parola con due sensi principali potrebbe essere rappresentata da una curva che connette due regioni dense nello spazio, dove le estremità della curva corrispondono ai sensi prototipici e i punti intermedi rappresentano usi ambigui o di transizione.

Un secondo approccio modella la polisemia attraverso distribuzioni di probabilità sullo spazio semantico. Invece di un singolo vettore v , una parola polisemica è rappresentata da una distribuzione $p(v|w)$, tipicamente una mistura di gaussiane $p(v|w) = \sum_k \pi_k N(v | \mu_k, \Sigma_k)$, dove k varia sui sensi della parola, π_k sono i pesi di mistura (che sommano a uno), μ_k sono i centroidi dei sensi, e Σ_k sono le matrici di covarianza che catturano l'incertezza o la variabilità di ciascun senso. Questo approccio permette di quantificare non solo dove si trovano i sensi nello spazio, ma anche la loro densità relativa e la loro dispersione.

La geometria delle distribuzioni di probabilità su varietà riemanniane, oggetto di studio della geometria dell'informazione, fornisce strumenti per confrontare e manipolare queste rappresentazioni probabilistiche. La metrica di Fisher sulla varietà delle distribuzioni gaussiane multivariate può essere usata per misurare la distanza tra rappresentazioni polisemiche. Per due distribuzioni gaussiane $N(\mu_1, \Sigma_1)$ e $N(\mu_2, \Sigma_2)$, esiste una formula esplicita per la distanza di Fisher che coinvolge sia le medie che le matrici di covarianza, fornendo una misura geometricamente naturale di quanto due rappresentazioni polisemiche differiscono.

Varietà di Sensi e Fibrati Semantici

Un framework matematico più sofisticato per la polisemia utilizza la teoria dei fibrati. Possiamo concepire uno spazio totale E dove ogni punto è una coppia (w, s) consistente di una forma lessicale w e uno specifico senso s di quella forma. Esiste una proiezione naturale $\pi: E \rightarrow W$ che mappa (w, s) alla forma w , e per ogni forma $w \in W$, la fibra $\pi^{-1}(w)$ è lo spazio dei sensi di w . La struttura di fibrato cattura l'idea che i sensi sono organizzati "sopra" le forme lessicali.

Se lo spazio base W è dotato di una topologia (per esempio, la topologia discreta dove ogni sottoinsieme è aperto), e ciascuna fibra $\pi^{-1}(w)$ è uno spazio topologico (per esempio, uno spazio metrico derivato dalla similarità semantica tra sensi), allora E eredita una topologia naturale. Sezioni di questo fibrato, ovvero funzioni $\sigma: W \rightarrow E$ tali che $\pi \circ \sigma = \text{id}_W$, corrispondono a scelte di un senso specifico per ogni parola, ovvero disambiguazioni globali.

Per parole polisemiche regolari, dove i pattern di polisemia si ripetono attraverso classi di parole (per esempio, la polisemia animale/carne che si applica a "pollo", "agnello", "maiale"), le fibre possono avere strutture simili. Questo può essere formalizzato attraverso l'idea di fibrati con gruppo di struttura: esiste un gruppo G che agisce sulle fibre, e le trivializzazioni locali sono compatibili con questa azione. Nel caso della polisemia regolare, G potrebbe essere un gruppo di trasformazioni semantiche (come la trasformazione da animale a carne) che si applica consistentemente attraverso diverse forme lessicali.

Gli embedding contestualizzati possono essere interpretati nel framework dei fibrati come segue. Per ogni contesto c , abbiamo una sezione σ_c che assegna a ogni parola w il suo

embedding contestualizzato $\phi(w, c)$. La collezione di tutte queste sezioni, al variare di c , popola lo spazio totale E . La struttura geometrica di E , in particolare la distribuzione dei punti nelle fibre, codifica la struttura polisemica del lessico. Parole monosemiche hanno fibre con bassa varianza, concentrate intorno a un singolo punto, mentre parole fortemente polisemiche hanno fibre con alta varianza, possibilmente multimodali.

Clustering e Disambiguazione nei Sensi

Data una collezione di embedding contestualizzati $\{\phi(w, c_i)\}$ per una parola w attraverso diversi contesti c_i , il problema della sense induction consiste nell'identificare i diversi sensi di w attraverso clustering. Algoritmi standard come k-means o clustering gerarchico possono essere applicati, ma la scelta della metrica e del numero di cluster è cruciale e non banale.

Un approccio sofisticato utilizza modelli di mistura gaussiana (GMM) con inferenza bayesiana. Assumiamo che gli embedding contestualizzati siano generati da una mistura $p(v) = \sum_k \pi_k N(v | \mu_k, \Sigma_k)$, e utilizziamo l'algoritmo Expectation-Maximization (EM) o metodi MCMC per inferire i parametri. Il passo E calcola le probabilità posteriori $p(k | v_i) \propto \pi_k N(v_i | \mu_k, \Sigma_k)$ che l' i -esimo embedding appartenga al k -esimo senso. Il passo M aggiorna i parametri: $\pi_k = (1/N) \sum_i p(k | v_i)$, $\mu_k = \sum_i p(k | v_i) v_i / \sum_i p(k | v_i)$, $\Sigma_k = \sum_i p(k | v_i) (v_i - \mu_k)(v_i - \mu_k)^T / \sum_i p(k | v_i)$.

Un vantaggio del GMM è che fornisce naturalmente probabilità di appartenenza soft piuttosto che assegnazioni hard, catturando così l'ambiguità intrinseca di alcuni usi. Inoltre, criteri come il Bayesian Information Criterion (BIC) o il Akaike Information Criterion (AIC) possono essere usati per selezionare automaticamente il numero di sensi, bilanciando fit dei dati e complessità del modello.

Approcci basati su grafi costruiscono un grafo di similarità dove i nodi sono gli embedding contestualizzati e gli archi connettono embedding sufficientemente simili. Algoritmi di community detection come Louvain o Infomap identificano gruppi densamente connessi che corrispondono a sensi. Questo approccio può catturare strutture più complesse rispetto al clustering basato su centroidi, come sensi con forme non sferiche o sensi gerarchicamente strutturati.

Transizioni Semantiche e Teoria delle Catastrofi

La teoria delle catastrofi, sviluppata da René Thom negli anni '60 e '70, studia come i sistemi dinamici possono subire cambiamenti qualitativi repentini in risposta a variazioni graduali dei parametri. Applicata alla semantica, questa teoria fornisce un framework per comprendere fenomeni come il cambiamento semantico, le transizioni tra sensi polisemici, e l'emergere di nuovi significati.

Consideriamo una famiglia di potenziali $V(x, \alpha)$, dove $x \in \mathbb{R}^n$ rappresenta una configurazione semantica (per esempio, la posizione nello spazio semantico) e $\alpha \in \mathbb{R}^m$ è un parametro di controllo (per esempio, il contesto, il tempo storico, o attributi pragmatici). Gli stati stabili del sistema corrispondono ai minimi locali di V , e l'evoluzione del sistema segue la dinamica del gradiente $dx/dt = -\nabla_x V(x, \alpha)$.

Al variare del parametro α , la struttura dei minimi locali può cambiare qualitativamente. Minimi possono apparire, scomparire, o fondersi. Quando due minimi si fondono, il sistema subisce una catastrofe, e una transizione repentina da uno stato all'altro diventa possibile. La teoria delle catastrofi classifica le catastrofi generiche in bassa dimensione attraverso forme normali standard: la catastrofe fold (piega), la cusp (cuspidale), la swallowtail (coda di rondine), e altre.

La catastrofe cusp, forse la più studiata, coinvolge un potenziale $V(x, \alpha, \beta) = x^4/4 + \alpha x^2/2 + \beta x$, dove (α, β) sono parametri di controllo bidimensionali. Per $\alpha > 0$, esiste un unico minimo stabile. Per $\alpha < 0$, possono esistere due minimi separati da un massimo locale. La transizione tra il regime unimodale e bimodale avviene sulla curva $\alpha = 0$, ma l'effettiva transizione che un sistema in evoluzione sperimenta dipende dal percorso nel piano (α, β) . Questo porta a isteresi: il sistema può "ricordare" stati precedenti e la transizione può avvenire a valori diversi di α a seconda della direzione del cambiamento.

Nel contesto semantico, i due minimi possono corrispondere a due sensi di una parola polisemica, e i parametri di controllo a variabili contestuali. La catastrofe cusp predice che, al variare del contesto, la parola può transitare bruscamente da un senso all'altro, e che esiste una regione di bistabilità dove entrambi i sensi sono accessibili. Questo modello è stato applicato per spiegare fenomeni come l'emergenza di nuovi sensi metaforici e la stabilizzazione di polisemie.

Dinamica Evolutiva degli Spazi Semantici

Il significato delle parole evolve nel tempo, e questa evoluzione può essere studiata attraverso la lente della dinamica su varietà. Consideriamo una famiglia di spazi semantici S_t parametrizzata dal tempo t , dove S_t rappresenta la configurazione dello spazio lessicale all'epoca t . L'evoluzione può essere descritta attraverso un'equazione differenziale sulla varietà degli spazi vettoriali, o equivalentemente attraverso flussi su gruppi di trasformazioni.

Un modello semplice assume che l'embedding $\phi_t(w)$ di una parola w evolva secondo $d\phi_t(w)/dt = F(\phi_t(w), t)$, dove F è un campo vettoriale che può dipendere esplicitamente dal tempo. Se F è derivato dal gradiente di una funzione obiettivo che cambia nel tempo, come quella che emerge dall'apprendimento continuo su corpora diacronici, l'evoluzione può essere vista come una discesa del gradiente su una superficie di energia che si deforma.

Le isometrie temporali, ovvero mappe $\phi_t \rightarrow \phi_s$ che preservano le strutture geometriche, permettono di allineare spazi semantici di epoche diverse. Dato due spazi S_t e S_s , una mappa lineare ortogonale $W: S_t \rightarrow S_s$ che minimizza $\sum_w \|\phi_t(w) - \phi_s(w)\|^2$ allinea ottimalmente i due spazi. Questa mappa può essere calcolata attraverso l'analisi di Procrustes: $W = UV^T$, dove U e V sono ottenuti dalla decomposizione SVD della matrice di correlazione crociata $\sum_w \phi_s(w) \phi_t(w)^T = U \Sigma V^T$.

Dopo l'allineamento, la distanza $\|\phi_t(w) - \phi_s(w)\|$ quantifica quanto il significato della parola w è cambiato tra i tempi t e s . Parole con grande distanza hanno subito spostamenti semantici significativi. L'analisi delle direzioni di questi spostamenti può rivelare tendenze sistematiche, come generalizzazioni, specializzazioni, ameliorazioni (miglioramento della connotazione), o pejorazioni (peggioramento della connotazione).

La teoria delle biforcazioni, strettamente legata alla teoria delle catastrofi, studia come le strutture qualitative delle soluzioni di equazioni differenziali cambiano al variare dei parametri. Nel contesto semantico, una biforcazione potrebbe corrispondere alla divisione di un significato originariamente unitario in due sensi distinti (biforcazione a forchetta), o alla nascita di un nuovo senso stabile (biforcazione transcritica). L'analisi locale vicino a punti di biforcazione utilizza forme normali e teoria del centro per caratterizzare la dinamica.

Spazi di Forme e Metriche di Deformazione

Lo spazio delle forme (shape space) di Kendall fornisce un framework per studiare configurazioni geometriche modulo trasformazioni di similarità (traslazioni, rotazioni, e scalature uniformi). Nel contesto semantico, possiamo considerare lo spazio delle configurazioni di un insieme fisso di parole modulo isometrie dello spazio ambiente. Due configurazioni sono considerate equivalenti se possono essere sovrapposte attraverso una rotazione e una scalatura.

Formalmente, dato un insieme di k parole con embeddings $\phi(w_1), \dots, \phi(w_k) \in \mathbb{R}^n$, la configurazione è rappresentata dalla matrice $X = [\phi(w_1), \dots, \phi(w_k)] \in \mathbb{R}^{n \times k}$. Lo spazio delle pre-forme è l'insieme di tali matrici modulo traslazioni, ottenuto centrando: $X_c = X - (1/k) X \mathbf{1} \mathbf{1}^T$, dove $\mathbf{1}$ è il vettore di tutti uno. Lo spazio delle forme è ottenuto quotizzando per rotazioni e scalature: $[X] = \{RX_c s : R \in SO(n), s > 0\}$.

La metrica di Procrustes su questo spazio misura la distanza minima tra due configurazioni dopo allineamento ottimale: $d([X], [Y]) = \inf_{R,s} \|X_c - sRY_c\|_F$, dove $\|\cdot\|_F$ è la norma di Frobenius. Questa metrica induce una geometria riemanniana sullo spazio delle forme, e geodetiche corrispondono a deformazioni ottimali che trasformano gradualmente una configurazione in un'altra.

Nel contesto semantico, l'analisi dello spazio delle forme può rivelare se la struttura relativa di un campo semantico è preservata attraverso lingue, domini, o epoche, anche se le scale assolute differiscono. Per esempio, le relazioni spaziali tra termini di colore potrebbero avere una struttura universale nello shape space, anche se le lingue lessicalizzano lo spettro cromatico in modo diverso.

Geometria Simplettica e Strutture Hamiltoniane

La geometria simplettica studia varietà dotate di una forma simplettica ω , una forma differenziale chiusa e non degenere. Mentre la geometria riemanniana è adatta per studiare "distanze" e "angoli", la geometria simplettica è naturale per studiare "flussi" e "conservazione". Sebbene meno comune nella semantica rispetto alla geometria riemanniana, la struttura simplettica può emergere in contesti dove il significato è visto come un sistema dinamico con quantità conservate.

Una forma simplettica in coordinate canoniche $(q, p) \in \mathbb{R}^{2n}$ è $\omega = \sum_i dq_i \wedge dp_i$. Data una funzione hamiltoniana $H: \mathbb{R}^{2n} \rightarrow \mathbb{R}$, il campo vettoriale hamiltoniano X_H è definito implicitamente da $\omega(X_H, \cdot) = dH$. Le equazioni di Hamilton sono $dq_i/dt = \partial H / \partial p_i$, $dp_i/dt = -\partial H / \partial q_i$. Il flusso hamiltoniano preserva la forma simplettica e la funzione H (energia).

Nel contesto semantico, se interpretiamo una metà delle coordinate come "posizioni" nello spazio semantico e l'altra metà come "momenti" o "velocità di cambiamento", una struttura hamiltoniana potrebbe descrivere l'evoluzione del significato sotto vincoli di conservazione. Per esempio, in un modello dove il significato evolve per minimizzare una funzione di costo mentre preserva certe quantità (come la distintività lessicale o l'informazione mutua con i referenti), la dinamica potrebbe essere naturalmente hamiltoniana.

Le trasformazioni canoniche sono simplettomorfismi, ovvero diffeomorfismi che preservano la forma simplettica. Queste trasformazioni formano un gruppo infinito-dimensionale e possono rappresentare cambiamenti di coordinate o di rappresentazione che preservano la struttura dinamica intrinseca. Nel contesto semantico, trasformazioni canoniche potrebbero corrispondere a riparametizzazioni dello spazio che preservano le relazioni dinamiche tra significati.

Informazione Quantistica e Spazi di Hilbert Semantici

Alcuni approcci alla semantica si ispirano al formalismo della meccanica quantistica, rappresentando significati come vettori in spazi di Hilbert complessi e composizione semantica come operazioni tensoriali o coinvolgimento quantistico (entanglement). Sebbene controverso e spesso metaforico, questo approccio ha prodotto modelli computazionali interessanti per fenomeni come l'interferenza semantica e la non commutatività della composizione.

Uno spazio di Hilbert H è uno spazio vettoriale complesso dotato di un prodotto scalare hermitiano $\langle \cdot, \cdot \rangle$ che è completo rispetto alla norma indotta. Stati quantistici sono rappresentati da vettori normalizzati $|\psi\rangle \in H$ con $\langle \psi | \psi \rangle = 1$, o più generalmente da operatori densità ρ che sono hermitiani, positivi semi-definiti, e hanno traccia uno. Il prodotto tensoriale $H_1 \otimes H_2$ di due spazi di Hilbert rappresenta sistemi composti.

Nella semantica quantistica, una parola potrebbe essere rappresentata da un operatore densità ρ_w che codifica non solo un significato medio ma anche l'incertezza e le correlazioni tra componenti semantiche. La composizione di due parole è modellata attraverso il prodotto tensoriale $\rho_{w_1} \otimes \rho_{w_2}$, possibilmente seguito da una misura o da un'operazione unitaria che seleziona o modula aspetti del significato composto.

Il fenomeno dell'interferenza quantistica è stato proposto per spiegare effetti di ordine nelle parole e ambiguità contestuali. Se due percorsi compositivi portano allo stesso stato finale, le loro ampiezze di probabilità possono interferire costruttivamente o distruttivamente, portando a probabilità finali che differiscono dalla semplice somma delle probabilità classiche. Questo è stato usato per modellare giudizi di similarità che violano proprietà classiche come la disuguaglianza triangolare in certi contesti.

L'entanglement quantistico, dove lo stato di un sistema composto non può essere fattorizzato come prodotto di stati dei sottosistemi, è stato proposto come modello per l'olismo semantico, l'idea che il significato di una composizione non sia riducibile ai significati delle parti. Stati entangled massimalmente, come gli stati di Bell, rappresentano la massima correlazione non classica e potrebbero modellare idiomi o costruzioni altamente non composizionali.

Geometria Kähleriana e Strutture Complesse

Una varietà kähleriana è una varietà differenziabile dotata simultaneamente di una struttura complessa, una metrica riemanniana, e una forma simplettica, in modo compatibile.

Specificamente, una struttura complessa è un campo tensoriale $J: TM \rightarrow TM$ con $J^2 = -I$ (dove I è l'identità). La compatibilità richiede che $g(JX, JY) = g(X, Y)$ (la metrica è hermitiana rispetto a J) e che $\omega(X, Y) = g(JX, Y)$ definisca una forma simplettica chiusa.

Le varietà kähleriane hanno proprietà geometriche molto ricche. Per esempio, l'operatore di Laplace-Beltrami si decompone come $\Delta = 2(\partial\bar{\partial}^* + \bar{\partial}\partial)$, dove ∂ e $\bar{\partial}$ sono operatori di Dolbeault. Le geodetiche in varietà kähleriane hanno proprietà speciali riguardo le curve olomorfe. La curvatura ha componenti che vanno oltre il tensore di Ricci, inclusa la forma di Ricci e la curvatura scalare olomorfa.

Nel contesto semantico, una struttura complessa potrebbe emergere se lo spazio semantico ammette una naturale decomposizione in parte "reale" e "immaginaria", dove la parte immaginaria potrebbe rappresentare aspetti connotativi, emotivi, o pragmatici del significato che interagiscono con il significato denotativo (reale) in modo strutturato. La compatibilità kähleriana garantirebbe che queste diverse facce del significato siano intimamente intrecciate in modo geometricamente coerente.

Gruppi di Lie e Simmetrie Semantiche

I gruppi di Lie sono varietà differenziabili dotate di una struttura di gruppo che è compatibile con la struttura differenziabile. Esempi fondamentali includono il gruppo lineare generale $GL(n, \mathbb{R})$, il gruppo ortogonale $O(n)$, il gruppo speciale ortogonale $SO(n)$, e il gruppo unitario $U(n)$. I gruppi di Lie agiscono naturalmente su varietà, e queste azioni catturano simmetrie.

Nel contesto semantico, certi gruppi di Lie possono agire sullo spazio semantico preservando strutture importanti. Per esempio, se lo spazio semantico è la sfera S^{n-1} , il gruppo $SO(n)$ agisce transitivamente su di essa per rotazioni. Sottogruppi di $SO(n)$ che fissano certe direzioni o iperpiani corrispondono a simmetrie che preservano certi attributi semantici.

Un'azione di un gruppo di Lie G su una varietà M è una mappa liscia $G \times M \rightarrow M$ che soddisfa proprietà di identità e composizione. Lo spazio delle orbite M/G , ottenuto identificando punti collegati dall'azione del gruppo, può avere una struttura di varietà (possibilmente con singolarità) e rappresenta l'insieme delle "configurazioni essenzialmente distinte" modulo la simmetria. Nel contesto semantico, M/G potrebbe rappresentare significati genuinamente diversi dopo aver quozientato per variazioni superficiali o ridondanti.

L'algebra di Lie $\mathfrak{g}$ di un gruppo di Lie G è lo spazio tangente all'identità, dotato del bracket di Lie $[\cdot, \cdot]$. Per matrici, il bracket di Lie è semplicemente il commutatore $[A, B] = AB - BA$. L'algebra di Lie linearizza la struttura del gruppo vicino all'identità e permette di studiare le proprietà infinitesimali del gruppo. Rappresentazioni dell'algebra di Lie, ovvero omomorfismi $\mathfrak{g} \rightarrow \mathfrak{gl}(V)$ dove V è uno spazio vettoriale, descrivono come il gruppo agisce su V attraverso trasformazioni infinitesimali.

Teoria dei Grafi Quantistici e Reti Semantiche

La teoria dei grafi fornisce un'altra lente per studiare strutture semantiche, specialmente quando le relazioni discrete tra parole sono enfatizzate rispetto alla struttura metrica continua. Un grafo semantico $G = (V, E)$ ha parole come vertici V e archi E che connettono parole semanticamente correlate. Diverse relazioni semantiche (sinonimia, antonimia, iperonimia, meronimia) possono corrispondere a diversi tipi di archi, portando a grafi multipli o multigraph.

La matrice di adiacenza A di un grafo non orientato ha $A_{ij} = 1$ se esiste un arco tra i vertici i e j , e 0 altrimenti. La matrice laplaciana del grafo è $L = D - A$, dove D è la matrice diagonale dei gradi. Gli autovalori e autovettori di L codificano proprietà spettrali del grafo. L'autovalore più piccolo è sempre 0, e la sua molteplicità è il numero di componenti connesse. Il secondo autovalore più piccolo, detto gap spettrale o connettività algebrica, misura quanto il grafo è connesso.

Gli embedding spettrali utilizzano gli autovettori della matrice laplaciana per assegnare coordinate ai vertici. L'embedding di Fiedler utilizza l'autovettore del secondo autovalore più piccolo per partizionare il grafo in due parti bilanciate. Embedding in dimensioni superiori utilizzano autovettori multipli, fornendo coordinate che tendono a posizionare vertici connessi vicini e vertici sconnessi lontani.

I random walk sul grafo semantico forniscono un altro metodo per derivare embeddings. La probabilità di transizione da i a j in un random walk è $P_{ij} = A_{ij} / d_i$, dove d_i è il grado del vertice i . Algoritmi come DeepWalk e Node2Vec generano sequenze di nodi attraverso random walks e poi applicano tecniche di embedding (come Skip-gram) a queste sequenze, trattandole come "frasi" dove i nodi giocano il ruolo di parole.

La diffusione su grafi, governata dall'equazione del calore $\partial u / \partial t = -Lu$, descrive come una distribuzione di probabilità $u(t)$ evolve nel tempo. La soluzione è $u(t) = e^{-tL} u(0)$, dove l'esponenziale della matrice è definito tramite la serie di potenze. Questo processo diffusivo cattura l'idea di propagazione di attivazione semantica attraverso una rete di concetti.

Applicazioni Avanzate, Prospettive Interdisciplinari e Sintesi Teorica

Geometria delle Rappresentazioni Multimodali

L'estensione degli spazi semantici al dominio multimodale, dove testo, immagini, audio e altre modalità sono rappresentati in uno spazio comune o in spazi accoppiati, introduce nuove complessità geometriche e topologiche. L'obiettivo fondamentale è costruire rappresentazioni dove oggetti semanticamente correlati, indipendentemente dalla modalità di origine, siano vicini nello spazio, permettendo compiti come il recupero cross-modale, la generazione caption, e la risposta a domande visive.

Un approccio consiste nel definire spazi separati per ciascuna modalità, V_{testo} e V_{immagine} , e poi apprendere mappe di allineamento $f: V_{\text{testo}} \rightarrow V_{\text{comune}}$ e $g: V_{\text{immagine}} \rightarrow V_{\text{comune}}$.

$V_{\text{immagine}} \rightarrow V_{\text{comune}}$ che proiettano entrambe le modalità in uno spazio comune. Queste mappe possono essere apprese minimizzando una funzione di perdita contrastiva che attrae rappresentazioni di coppie corrispondenti (come una frase e l'immagine che descrive) mentre respinge rappresentazioni di coppie non corrispondenti. La perdita tripla è una scelta comune: $L = \max(0, m + d(f(t), g(i^*)) - d(f(t), g(i^-)))$, dove (t, i^*) è una coppia positiva, i^- è un'immagine negativa, e m è un margine.

La geometria dello spazio comune eredita proprietà sia dallo spazio testuale che da quello visivo, ma emergono anche nuove strutture. Per esempio, oggetti visivi che condividono attributi verbali (come "rosso" o "rotondo") possono formare cluster che riflettono la tassonomia lessicale, mentre oggetti con apparenze visive simili ma nomi diversi formano cluster ortogonali. La struttura risultante può essere altamente non convessa e frammentata.

Le varietà prodotto $V_{\text{testo}} \times V_{\text{immagine}}$, dotate della metrica prodotto, offrono un framework alternativo dove ciascun elemento ha coordinate in entrambi gli spazi. La proiezione su ciascun fattore permette di recuperare la rappresentazione modale singola, mentre la rappresentazione congiunta cattura correlazioni multimodali. Operazioni tensoriali sul prodotto permettono di catturare interazioni complesse, come l'attenzione cross-modale dove regioni dell'immagine sono ponderate in funzione del contenuto testuale.

Teoria della Categorizzazione e Spazi Concettuali

La teoria dei prototipi di Rosch propone che i concetti siano organizzati intorno a membri prototipici centrali, con membri periferici che gradano in tipicità. Geometricamente, questo si traduce in una struttura dove ogni categoria è rappresentata da una regione dello spazio con un centroide prototipico e una funzione di appartenenza graduale che decresce con la distanza dal prototipo. La funzione di appartenenza può essere modellata come una gaussiana $\mu_C(x) = \exp(-d(x, c_C)^2/(2\sigma^2))$, dove c_C è il prototipo della categoria C e σ controlla la "larghezza" della categoria.

La teoria degli spazi concettuali di Gärdenfors formalizza questa intuizione assumendo che i concetti siano regioni convesse in spazi geometrici strutturati da dimensioni qualitative. Una dimensione qualitativa è un'ordinazione dei valori lungo un continuum, come la temperatura, il peso, o la brillantezza cromatica. Lo spazio concettuale è il prodotto cartesiano di queste dimensioni, e la convessità riflette l'assunzione naturalistica che se due oggetti appartengono a una categoria, anche gli oggetti intermedi (in senso geometrico) dovrebbero appartenervi.

La convessità ha implicazioni topologiche profonde. Insiemi convessi in spazi euclidei sono contrattili, ovvero omotopicamente equivalenti a un punto, il che significa che sono topologicamente semplici senza "buchi". Questo si accorda con l'intuizione che i concetti naturali dovrebbero essere coerenti e unificati. Tuttavia, concetti complessi o disgiunti (come "essere un numero primo" o "essere una proprietà emergenziale") possono violare la convessità, richiedendo strutture più ricche come unioni di regioni convesse.

Le distanze di Hausdorff tra regioni concettuali forniscono una metrica sullo spazio delle categorie stesse. Data due regioni compatte A e B , la distanza di Hausdorff è $d_H(A, B) = \max(\sup_{a \in A} \inf_{b \in B} d(a, b), \sup_{b \in B} \inf_{a \in A} d(a, b))$. Questa metrica quantifica

quanto due categorie sono "lontane" come insiemi, considerando sia la massima distanza di un punto di A dal punto più vicino di B, sia viceversa.

Geometria della Metafora e Spazi Amalgamati

La metafora concettuale, come teorizzata da Lakoff e Johnson, coinvolge la mappatura strutturata di un dominio sorgente su un dominio target. Geometricamente, questo può essere modellato come una mappa $f: S \rightarrow T$ tra sottospazi o sottoinsiemi dello spazio semantico, dove S è la regione del dominio sorgente e T del dominio target. La mappa non è arbitraria ma preserva certe relazioni strutturali, una proprietà che può essere formalizzata attraverso il concetto di omomorfismo parziale.

La teoria del blending concettuale di Fauconnier e Turner propone che la comprensione metaforica coinvolga la costruzione di uno spazio amalgamato (blend) che eredita strutture da entrambi i domini input ma può anche sviluppare strutture emergenti. Geometricamente, il blend può essere visto come un fibrato su uno spazio base che parametrizza il "grado di mescolamento", o come una sottovarietà che interpola tra le varietà dei domini sorgente e target.

Consideriamo una formalizzazione attraverso la teoria delle categorie. Abbiamo spazi (o categorie) S e T , uno spazio generico G che cattura la struttura comune astratta, e mappe $\varphi_S: G \rightarrow S$ e $\varphi_T: G \rightarrow T$ che proiettano la struttura generica sui domini specifici. Lo spazio amalgamato B è costruito come colimite (pushout) nel diagramma $S \leftarrow G \rightarrow T$, ovvero il più piccolo spazio che riceve mappe da S e T consistenti con le proiezioni da G .

In termini geometrici, se G, S, T sono varietà e le mappe sono immersioni, il blend B può essere costruito come la varietà che risulta dall'incollamento di S e T lungo G . La geometria di B riflette sia la geometria di S che quella di T , e le geodetiche in B possono "saltare" tra le porzioni ereditate da S e T , rappresentando percorsi concettuali che sfruttano la metafora per navigare tra domini.

Geometria dell'Informazione e Apprendimento Statistico

La geometria dell'informazione di Amari fornisce una cornice teorica per studiare spazi di modelli probabilistici attraverso lenti geometriche. Come menzionato precedentemente, la metrica di Fisher $g_{ij}(\theta) = E[\partial \log p(x|\theta)/\partial \theta_i \partial \log p(x|\theta)/\partial \theta_j]$ definisce una metrica riemanniana sullo spazio dei parametri. Oltre alla metrica, la geometria dell'informazione introduce due connessioni affini duali, la connessione esponenziale (e) e la connessione mistura (m), che generalizzano la connessione di Levi-Civita.

La dualità tra queste connessioni è espressa dalla relazione $\nabla^{\{e\}}_X Y + \nabla^{\{m\}}_X Y = X[g(\cdot, Y)]$, che generalizza la formula per la derivata covariante di Levi-Civita. Le geodetiche rispetto a queste connessioni hanno interpretazioni statistiche: le e -geodetiche sono curve esponenziali della forma $p(x|\theta(t)) = p(x|\theta_0) \exp(t \eta \cdot T(x) - \psi(t))$, dove $T(x)$ è una statistica sufficiente, mentre le m -geodetiche sono misture lineari $p(x|\theta(t)) = (1-t)p(x|\theta_0) + t p(x|\theta_1)$.

La curvatura rispetto a queste connessioni può essere diversa da zero anche quando la varietà è localmente piatta in senso riemanniano, riflettendo complessità intrinseche della famiglia di distribuzioni. Per famiglie esponenziali, la connessione esponenziale è piatta, e

esistono coordinate (dette coordinate naturali) dove la connessione sparisce. Questo ha implicazioni per l'efficienza statistica: stimatori che sono efficienti in un sistema di coordinate rimangono efficienti sotto cambiamenti di coordinate che preservano la struttura affine.

Nell'apprendimento di embeddings, la funzione di perdita può essere vista come una funzione sulla varietà statistica delle distribuzioni indotte dagli embeddings. La discesa del gradiente naturale, che utilizza il gradiente riemanniano rispetto alla metrica di Fisher piuttosto che il gradiente euclideo, può convergere più rapidamente poiché tiene conto della geometria intrinseca dello spazio dei parametri. Il gradiente naturale è dato da $\tilde{\nabla} L = G^{-1} \nabla L$, dove G è la matrice di informazione di Fisher e ∇L è il gradiente euclideo.

Teoria dell'Omotopia e Invarianti Algebrici

La teoria dell'omotopia studia le varietà e gli spazi topologici attraverso continue deformazioni. Due spazi X e Y sono omotopicamente equivalenti se esistono mappe continue $f: X \rightarrow Y$ e $g: Y \rightarrow X$ tali che $g \circ f$ è omotopa all'identità su X e $f \circ g$ è omotopa all'identità su Y . L'equivalenza omotopica è più debole dell'omeomorfismo (equivalenza topologica) ma cattura l'essenza topologica "elastica" degli spazi.

I gruppi di omotopia $\pi_n(X, x_0)$ sono invarianti algebrici fondamentali. Il gruppo fondamentale $\pi_1(X, x_0)$ classifica le classi di omotopia di loop basati in x_0 , ovvero curve continue $\gamma: [0, 1] \rightarrow X$ con $\gamma(0) = \gamma(1) = x_0$. Per $n \geq 2$, il gruppo $\pi_n(X, x_0)$ classifica mappe dalla n -sfera S^n in X , modulo omotopia. Per spazi semplicemente connessi ($\pi_1 = 0$), come la sfera, la struttura omotopica è ancora molto ricca: $\pi_2(S^2) = \mathbb{Z}$, $\pi_3(S^2) = \mathbb{Z}$, $\pi_4(S^2) = \mathbb{Z}/2\mathbb{Z}$, e la sequenza continua in modo complesso.

Nel contesto semantico, il gruppo fondamentale potrebbe catturare l'esistenza di "loop semantici" irriducibili, ovvero catene cicliche di relazioni tra parole che non possono essere contratte a un punto. Per esempio, se seguiamo una catena di sinonimie o associazioni che ritorna alla parola iniziale ma attraversa significati intermedi genuinamente diversi, questo costituisce un generatore non banale del gruppo fondamentale.

La teoria dell'omotopia stabile studia i gruppi di omotopia nel limite di dimensione alta, $\pi_n(S^k)$ per $k \rightarrow \infty$ con $n-k$ fissato. Questi gruppi stabilizzati hanno una struttura più regolare e sono legati alla K -teoria e alla teoria dei cobordismi. Sebbene altamente astratta, la teoria dell'omotopia stabile potrebbe avere applicazioni nella comprensione delle proprietà asintotiche degli spazi semantici in dimensione molto alta.

Teorie di Gauge e Connessioni su Fibrati

Le teorie di gauge in fisica descrivono interazioni fondamentali attraverso connessioni su fibrati principali. Traslando queste idee al contesto semantico, possiamo immaginare che diverse "fasi" o "prospettive" sul significato siano legate da trasformazioni di gauge, e che la fisica del significato sia invariante sotto tali trasformazioni.

Un fibrato principale $P(M, G)$ consiste di uno spazio totale P , uno spazio base M , e un gruppo di struttura G che agisce liberamente e transitivamente sulle fibre. Una connessione su P è un campo di sottospazi orizzontali negli spazi tangenti a P , complementari ai

sottospazi verticali (tangenti alle fibre). La connessione permette di "sollevare" curve in M a curve orizzontali in P , fornendo un modo di trasportare informazioni di gauge lungo percorsi.

La curvatura di una connessione misura il difetto di integrabilità dei sottospazi orizzontali. È un campo tensoriale che assume valori nell'algebra di Lie di G . In fisica, la curvatura corrisponde al campo di forza: in elettromagnetismo, il campo elettromagnetico è la curvatura di una connessione $U(1)$. Nel contesto semantico, la curvatura potrebbe rappresentare l'impossibilità di definire globalmente e consistentemente certi attributi semantici, richiedendo invece scelte locali (gauge) che sono legate da trasformazioni.

Le trasformazioni di gauge sono automorfismi del fibrato che preservano la struttura di fibrato. Localmente, corrispondono a mappe $M \rightarrow G$. La richiesta che le leggi fisiche (o, nel nostro caso, le relazioni semantiche) siano invarianti sotto trasformazioni di gauge impone vincoli severi sulla forma ammissibile di tali leggi, un principio estremamente fruttuoso in fisica che potrebbe avere analogie semantiche.

Geometria Algebrica e Varietà Affini

La geometria algebrica studia oggetti geometrici definiti da equazioni polinomiali. Una varietà algebrica affine è l'insieme delle soluzioni di un sistema di equazioni polinomiali in uno spazio affine. Sebbene gli spazi semantici siano tipicamente definiti attraverso apprendimento statistico piuttosto che equazioni esplicite, la prospettiva algebrica può emergere quando si impongono vincoli strutturali.

Per esempio, supponiamo di volere che certe relazioni semantiche siano catturate da equazioni polinomiali negli embeddings. Una relazione come "x è l'opposto di y" potrebbe essere modellata da $v_x + v_y = 0$, un'equazione lineare. Relazioni più complesse potrebbero richiedere polinomi di grado superiore. L'insieme di tutte le triple (v_x, v_y, v_z) che soddisfano una relazione ternaria polinomiale forma una varietà algebrica in $\mathbb{R}^{3n}$.

La dimensione di una varietà algebrica è un concetto fondamentale. Per una varietà definita da k equazioni indipendenti in $\mathbb{R}^n$, la dimensione è generalmente $n - k$. Varietà di dimensione uno sono curve, dimensione due sono superfici, e così via. Nel contesto semantico, vincoli algebrici che emergono da regolarità linguistiche riducono la dimensione effettiva dello spazio, concentrando i vettori di parole su varietà di dimensione inferiore.

Gli ideali polinomiali e le basi di Gröbner forniscono strumenti algoritmici per studiare varietà algebriche. Data un insieme di polinomi $\{f_1, \dots, f_k\}$, l'ideale $I = \langle f_1, \dots, f_k \rangle$ consiste di tutte le combinazioni polinomiali $\sum_i g_i f_i$. La varietà associata è $V(I) = \{x : f(x) = 0 \text{ per ogni } f \in I\}$. Calcolare una base di Gröbner per I fornisce una rappresentazione canonica che facilita la risoluzione di sistemi e il test di appartenenza.

Ottimizzazione su Varietà e Geometria Computazionale

L'ottimizzazione riemanniana è l'estensione dei metodi di ottimizzazione classica a varietà. L'idea fondamentale è di definire la discesa del gradiente e altri algoritmi iterativi in modo che rispettino la geometria della varietà. Data una funzione obiettivo $f: M \rightarrow \mathbb{R}$ su una varietà M , il gradiente riemanniano $\text{grad } f$ è l'unico campo vettoriale tale che $g(\text{grad } f, v) = df(v)$ per ogni vettore tangente v .

L'aggiornamento iterativo in discesa del gradiente riemanniano è $x_{k+1} = \text{Retr}_{x_k}(-\alpha_k \text{grad } f(x_k))$, dove $\text{Retr}: TM \rightarrow M$ è una retrazione, una mappa che approssima la mappa esponenziale. La mappa esponenziale $\exp_x: T_x M \rightarrow M$ è definita come $\exp_x(v) = \gamma_v(1)$, dove γ_v è la geodetica con $\gamma_v(0) = x$ e $\gamma'_v(0) = v$. Per varietà complesse, la mappa esponenziale può essere costosa da calcolare, e retrazioni più semplici come la proiezione sono preferite.

Per la sfera S^{n-1} , una retrazione semplice è $\text{Retr}_x(v) = (x + v)/\|x + v\|$, che proietta radialmente sulla sfera. Per lo spazio di Stiefel $\text{St}(n, k)$ delle matrici $n \times k$ ortogonali, retrazioni basate sulla fattorizzazione QR sono efficaci. Per varietà di Grassmann, che parametrizzano k -piani in $\mathbb{R}^n$, esistono retrazioni basate sulla decomposizione ai valori singolari.

I metodi di secondo ordine, come il metodo di Newton riemanniano, utilizzano l'Hessiana riemanniana per ottenere convergenza quadratica. L'Hessiana riemanniana $\text{Hess } f$ di una funzione f è un operatore che mappa vettori tangenti in vettori tangenti: $\text{Hess } f(v) = \nabla_v \text{grad } f$, dove ∇ è la derivata covariante. Il passo di Newton è $x_{k+1} = \text{Retr}_{x_k}(-\alpha_k (\text{Hess } f)^{-1} \text{grad } f(x_k))$. Questo richiede la risoluzione di un sistema lineare tangente a ogni iterazione.

Applicazioni alla Traduzione Automatica e Allineamento Multilingue

Gli spazi semantici multilingui cercano di rappresentare parole di lingue diverse in uno spazio comune, permettendo la traduzione e il trasferimento cross-linguale. Un approccio consiste nell'apprendere embeddings separatamente per ciascuna lingua e poi trovare una trasformazione (tipicamente ortogonale) che allinea gli spazi. Data una lista di coppie di parole tradotte $\{(w^s_i, w^t_i)\}$, si cerca una matrice ortogonale W che minimizza $\sum_i \|W v^s_i - v^t_i\|^2$.

La soluzione ottimale può essere ottenuta attraverso l'analisi di Procrustes, come menzionato precedentemente. Tuttavia, questa assume che le geometrie delle due lingue siano identiche modulo rotazione, un'assunzione che può essere violata a causa di differenze strutturali tra lingue. Approcci più sofisticati permettono trasformazioni non lineari o utilizzano adversarial training per apprendere mappe senza supervisione.

La teoria delle corrispondenze ottimali (optimal transport) fornisce un framework alternativo. Data distribuzioni di parole in due lingue, si cerca una mappa di trasporto che muove la massa di probabilità dalla distribuzione sorgente alla distribuzione target minimizzando un costo. La distanza di Wasserstein, già menzionata, quantifica questo costo minimo. Algoritmi efficienti per il calcolo del trasporto ottimale, come l'algoritmo di Sinkhorn, sono stati applicati all'allineamento di embeddings multilingui.

Le varietà prodotto $M_1 \times M_2 \times \dots \times M_k$, dove M_i è lo spazio semantico della lingua i , forniscono una rappresentazione unificata dove ogni concetto è una tupla di vettori in lingue diverse. La proiezione su ciascun fattore recupera l'embedding monolingue. Vincoli di accoppiamento o penalizzazioni possono essere imposti per incoraggiare coerenza semantica tra lingue.

Neuroscienze Computazionali e Rappresentazioni Cerebrali

L'interfaccia tra spazi semantici computazionali e rappresentazioni neurali nel cervello è un'area di ricerca attiva. Tecniche di neuroimaging come la risonanza magnetica funzionale (fMRI) misurano l'attività cerebrale mentre i soggetti processano stimoli linguistici. L'attività in ciascun voxel (volume element) del cervello può essere vista come una variabile dipendente, e gli embeddings semantici delle parole come predittori, permettendo la costruzione di modelli di encoding.

Un modello di encoding predice l'attività cerebrale $y \in \mathbb{R}^V$ (dove V è il numero di voxel) dato un embedding semantico $x \in \mathbb{R}^n$ attraverso una funzione $f: y = f(x) + \epsilon$. Modelli lineari $f(x) = Wx$, dove W è una matrice $V \times n$, sono comuni per la loro semplicità e interpretabilità. Le righe di W rappresentano "recettori semantici" per ciascun voxel, indicando quali dimensioni semantiche attivano quella regione cerebrale.

I modelli di decoding invertono la direzione: predicono proprietà dello stimolo (come quale parola è stata presentata) dall'attività cerebrale. Questo è rilevante per interfacce cervello-computer e per testare ipotesi sulla rappresentazione neurale. Se due parole sono vicine nello spazio semantico computazionale, ci si aspetta che producano pattern di attività cerebrale simili, una predizione testabile.

L'analisi di similarità rappresentazionale (RSA) confronta matrici di dissimilarità da diverse fonti (comportamentali, neurali, computazionali). Data una matrice di dissimilarità semantica D^{sem} , dove D^{sem}_{ij} misura la distanza tra parole i e j nello spazio semantico, e una matrice di dissimilarità neurale D^{neural} , dove D^{neural}_{ij} misura la distanza tra pattern di attività evocati dalle parole i e j , la correlazione tra D^{sem} e D^{neural} quantifica quanto il modello semantico cattura la struttura rappresentazionale neurale.

Filosofia del Significato e Ontologie Formali

La semantica formale in filosofia del linguaggio, sviluppata da Montague, Davidson, e altri, tratta il significato attraverso strutture logiche e teoria dei modelli. Il significato di una frase è spesso identificato con le sue condizioni di verità, ovvero le situazioni possibili in cui sarebbe vera. Gli spazi semantici geometrici offrono una prospettiva alternativa, più operativa e basata sull'uso, che si allinea con la filosofia del linguaggio ordinario di Wittgenstein.

Le ontologie formali, come quelle utilizzate nel Semantic Web e nell'ingegneria della conoscenza, strutturano i concetti in gerarchie di classi e proprietà. OWL (Web Ontology Language) permette di definire classi, sottoclassi, proprietà, restrizioni, e regole di inferenza. La relazione tra ontologie formali e spazi semantici geometrici è complessa: le ontologie forniscono strutture discrete e logiche, mentre gli spazi geometrici forniscono strutture continue e graduali.

Un ponte tra i due approcci può essere costruito attraverso fuzzy set theory, dove l'appartenenza a una classe non è binaria ma graduale. Un concetto fuzzy è caratterizzato da una funzione di appartenenza $\mu: X \rightarrow [0, 1]$ che assegna a ogni elemento x un grado di appartenenza. Nello spazio semantico, $\mu_C(v)$ potrebbe essere definito come una funzione della distanza di v da un prototipo di categoria, come menzionato precedentemente.

La teoria dei modelli categorica studia categorie di modelli e funtori tra esse, fornendo un framework astratto per confrontare diverse semantiche. Questo approccio ultra-astratto è rilevante quando si cercano relazioni profonde tra diversi sistemi di rappresentazione semantica, come embedding neurali, logiche descrittive, e reti semantiche.

Limiti e Frontiere della Geometrizzazione Semantica

Nonostante i successi degli spazi semantici, esistono limitazioni fondamentali. La linearità intrinseca degli spazi vettoriali impone vincoli: relazioni semantiche che richiedono strutture non lineari complesse potrebbero non essere catturabili fedelmente. La composizionalità, benché modellabile attraverso tensori e reti neurali, rimane parzialmente compresa, specialmente per costruzioni sintattiche complesse e idiomi.

La sensibilità al contesto e la pragmatica del linguaggio sfidano i modelli puramente geometrici. Il significato non è determinato solo dalla forma lessicale e dalla sintassi, ma anche da fattori contestuali, presupposizioni, implicature, e intenzionalità del parlante. Modelli contestualizzati catturano parzialmente questi aspetti, ma una teoria completa richiederebbe l'integrazione con modelli di interazione comunicativa e teoria degli atti linguistici.

La questione della interpretabilità è cruciale. Gli spazi ad alta dimensione sono difficili da visualizzare e comprendere. Mentre tecniche di riduzione dimensionale come t-SNE e UMAP forniscono proiezioni bidimensionali, queste inevitabilmente distorcono la geometria originale. Metodi per estrarre direzioni semantiche interpretabili (come "maschile-femminile" o "positivo-negativo") attraverso manipolazioni algebriche forniscono intuizioni, ma la maggior parte delle dimensioni rimane opaca.

La questione ontologica fondamentale persiste: gli spazi semantici sono strutture che scopriamo o che costruiamo? Riflettono la realtà del significato mentale o sono artefatti utili ma arbitrari del nostro approccio computazionale? Queste domande filosofiche profonde non hanno risposte definitive, ma la continua convergenza tra modelli geometrici, evidenze comportamentali, e correlati neurali suggerisce che c'è una realtà strutturale che viene catturata, anche se imperfettamente.

Direzioni Future e Prospettive Interdisciplinari

L'integrazione di geometrie non euclidee più esotiche, come geometrie sub-riemanniane, spazi di Finsler, e varietà con bordo o con singolarità, potrebbe catturare aspetti del significato finora trascurati. Le geometrie sub-riemanniane, dove il movimento è vincolato a sottospazi degli spazi tangenti, potrebbero modellare vincoli sulle trasformazioni semantiche ammissibili. Gli spazi di Finsler, dove la nozione di lunghezza dipende dalla direzione, potrebbero catturare asimmetrie nelle relazioni semantiche.

L'applicazione di tecniche della teoria delle categorie superiori, come ∞ -categorie e teoria dell'omotopia derivata, potrebbe fornire framework per studiare relazioni semantiche di ordine arbitrario e composizioni iterate in modo sistematico e astratto. La teoria dei topos, che generalizza la teoria degli insiemi e la logica classica, potrebbe offrire fondamenti alternativi per la semantica geometrica.

L'integrazione con l'intelligenza artificiale causale e l'apprendimento di rappresentazioni causali potrebbe estendere gli spazi semantici per catturare non solo correlazioni ma anche relazioni causali tra concetti. Questo richiederebbe l'incorporazione di strutture di grafo causale o equazioni strutturali negli embeddings, possibilmente attraverso varietà dotate di connessioni che codificano flussi causali.

Le applicazioni alla linguistica storica e all'evoluzione del linguaggio potrebbero beneficiare di modelli geometrici dinamici che tracciano le traiettorie semantiche attraverso secoli. L'analisi comparativa degli spazi semantici di lingue tipologicamente diverse potrebbe rivelare universali linguistici o vincoli cognitivi sulla struttura del significato. La collaborazione con antropologia cognitiva e psicolinguistica potrebbe validare o sfidare le predizioni dei modelli geometrici attraverso studi cross-culturali.

Conclusione: Verso una Teoria Unificata della Semantica Geometrica

La geometrizzazione del significato rappresenta uno dei programmi di ricerca più ambiziosi e interdisciplinari nella scienza cognitiva contemporanea. Attraverso l'integrazione di linguistica, matematica, informatica, neuroscienze e filosofia, stiamo costruendo una comprensione del significato come fenomeno fondamentalmente spaziale e strutturato, emergente dall'interazione tra regolarità statistiche del linguaggio, vincoli cognitivi, e geometrie sottostanti.

La potenza degli spazi semantici risiede nella loro capacità di rendere il significato computabile, misurabile, e visualizzabile, trasformando l'astratto in concreto. Le operazioni algebriche su vettori riflettono relazioni semantiche; le metriche quantificano similarità; le varietà differenziabili catturano strutture continue e fluide; le topologie rivelano connessioni globali; i tensori modellano composizioni complesse; e le geometrie non euclidee accomodano gerarchie e asimmetrie.

Tuttavia, questa geometrizzazione non è una riduzione meccanicista. Il significato mantiene la sua ricchezza, sfumatura, e dipendenza dal contesto. Gli spazi semantici sono strumenti potenti, ma non esaustivi. Essi illuminano aspetti del significato precedentemente inaccessibili all'analisi formale, ma lasciano spazio per fenomeni semantici che sfuggono alla cattura geometrica.

La frontiera della ricerca si sposta continuamente verso rappresentazioni più ricche, geometrie più espressive, e integrazioni più profonde con altre modalità di comprensione del significato. La teoria delle varietà, la topologia algebrica, la geometria differenziale, l'analisi funzionale, e la teoria delle categorie forniscono un arsenale matematico ineshausto. La linguistica teorica, la filosofia del linguaggio, e la psicologia cognitiva forniscono intuizioni e vincoli fenomenologici. Le neuroscienze forniscono ancoraggi empirici e vincoli biologici.

In definitiva, la topologia e la geometria degli spazi semantici non sono solo tecniche computazionali, ma costituiscono un nuovo linguaggio per pensare il significato, un framework concettuale che permette domande precedentemente non formulabili e risposte precedentemente inaccessibili.